

Social Virtue Epistemology

This collection of 19 chapters, all appearing in print here for the first time and written by an international team of established and emerging scholars, explores the place of intellectual virtues and vices in a social world. Relevant virtues include open-mindedness, curiosity, intellectual courage, diligence in inquiry, and the like. Relevant vices include dogmatism, need for immediate certainty, and gullibility and the like.

The chapters are divided into four key sections: Foundational Issues; Individual Virtues; Collective Virtues; and Methods and Measurements. And the chapters explore the most salient questions in these areas of research, including: How are individual intellectual virtues and vices affected by their social contexts? Does being in touch with other open-minded people make us more open-minded? Conversely, does connection to other dogmatic people make us more dogmatic? Can groups possess virtues and vices distinct from those of their members? For instance, could a group of dogmatic individuals operate in an open-minded way despite the vices of its members?

Each chapter receives commentary from two other authors in the volume, and each original author then replies to these commentaries. Together, the authors form part of a collective conversation about how we can know about what we know. In so doing, they not only theorize but enact social virtue epistemology.

Mark Alfano is Associate Professor of Philosophy at Macquarie University. In 2019, he published *Nietzsche's Moral Psychology* (Cambridge UP). His papers have won awards from the *Philosopher's Annual* (2018) and *Peritia* (2019).

Colin Klein is Professor of Philosophy at the Australian National University. He is the author of *What the Body Commands: The Imperative Theory of Pain* (MIT Press, 2015).

Jeroen de Ridder is Associate Professor of Philosophy at Vrije Universiteit Amsterdam and Professor by special appointment of Christian Philosophy at the University of Groningen. His research is in social and political epistemology, and in 2021 he co-edited *The Routledge Handbook of Political Epistemology*.

T&F Proofs – Not for Distribution

Social Virtue Epistemology

Edited by
Mark Alfano, Colin Klein,
and Jeroen de Ridder

 **Routledge**
Taylor & Francis Group
NEW YORK AND LONDON

First published 2022
by Routledge
605 Third Avenue, New York, NY 10158

and by Routledge
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2022 Taylor & Francis

The right of Mark Alfano, Colin Klein, and Jeroen de Ridder to be identified as the authors of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data
A catalog record for this title has been requested

ISBN: 978-0-367-40764-3 (hbk)
ISBN: 978-1-032-29120-8 (pbk)
ISBN: 978-0-367-80895-2 (ebk)

DOI: 10.4324/9780367808952

Typeset in Sabon
by codeMantra

Contents

<i>List of Figures</i>	xi
<i>List of Tables</i>	xiii
<i>Notes on Contributors</i>	xv
<i>Acknowledgements</i>	xvii
Introduction: A Research Program for Social Virtue Epistemology	1
MARK ALFANO, COLIN KLEIN, AND JEROEN DE RIDDER	
PART I	
Foundational Issues	13
1 Interactionism, Debiasing, and the Division of Epistemic Labour	15
STEVEN BLAND	
1b Commentary from Neil Levy	39
1c Commentary from Michel Croce and Duncan Pritchard	42
1d Steven Bland's Response to Commentaries	45
2 Attunement: On the Cognitive Virtues of Attention	48
GEORGI GARDINER	
2b Commentary from J. Adam Carter	73
2c Commentary from S. Goldberg	77
2d Georgi Gardiner's Response to Commentaries	80

vi *Contents*

3 From Vice Epistemology to Critical Character Epistemology IAN JAMES KIDD	84
3b Commentary from Heather Battaly	103
3c Commentary from Georgi Gardiner	106
3d Ian James Kidd's Response to Commentaries	110
4 Narrowing the Scope of Virtue Epistemology NEIL LEVY	113
4b Commentary from Steven Bland	131
4c Commentary from Quassim Cassam	134
4d Neil Levy's Response to Commentaries	137
5 Mindshaping and Intellectual Virtues ALESSANDRA TANESINI	140
5b Commentary from Ian James Kidd	161
5c Commentary from Thi Nguyen	165
5d Alessandra Tanesini's Response to Commentaries	167
PART II Individual Virtues and Vices	171
6 The Vices and Virtues of Extremism QUASSIM CASSAM	173
6b Commentary from Barend de Rooij & Boudewijn de Bruin	192
6c Commentary from Marco Meyer	195
6d Quassim Cassam's Response to Commentaries	198

7	Expectations of Expertise: Boot-Strapping in Social Epistemology	201
	SANFORD C. GOLDBERG	
7b	Commentary from Heidi Grasswick	223
7c	Commentary from Erik J. Olsson	227
7d	Sanford C. Goldberg's Response to Commentaries	232
8	Fake News, Conspiracy Theorizing, and Intellectual Vice	236
	MARCO MEYER AND MARK ALFANO	
8b	Commentary from Quassim Cassam	260
8c	Commentary from Colin Klein	263
8d	Marco Meyer and Mark Alfano's Response to Commentaries	266
9	Playfulness versus Epistemic Traps	269
	C. THI NGUYEN	
9b	Commentary from Ian James Kidd	291
9c	Commentary from Lani Watson	294
9d	C. Thi Nguyen's Response to Commentaries	298
PART III		
Collective Virtues and Vices		301
10	Solidarity: Virtue or Vice?	303
	HEATHER BATTALY	
10b	Commentary from T. Ryan Byerly	325
10c	Commentary from Duncan Pritchard	329
10d	Heather Battaly's Response to Commentaries	332

viii *Contents*

11	Collective (Telic) Virtue Epistemology J. ADAM CARTER	335
11b	Commentary from Jeroen de Ridder	357
11c	Commentary from S. Kate Devitt	360
11d	J. Adam Carter's Response to Commentaries	363
12	Three Models for Collective Intellectual Virtues JEROEN DE RIDDER	367
12b	Commentary from S. Kate Devitt	386
12c	Commentary from Heidi Grasswick	389
12d	Jeroen de Ridder's Response to Commentaries	393
13	Real-Life Collective Epistemic Virtue and Vice BAREND DE ROOIJ AND BOUDEWIJN DE BRUIN	396
13b	Commentary from Steven Bland	415
13c	Commentary from Neil Levy	418
13d	Barend de Rooij and Boudewijn de Bruin's Response to Commentaries	421
14	The Social Virtue of Questioning: A Genealogical Account LANI WATSON	424
14b	Commentary from J. Adam Carter	442
14c	Commentary from S. Goldberg	445
14d	Lani Watson's Response to Commentaries	448

PART IV	
Methods and Measurements	451
15 An Interdisciplinary Methodology for Studying Collective Intellectual Character Traits	453
T. RYAN BYERLY	
15b Commentary from Heather Battaly	470
15c Commentary from Marco Meyer	473
15d T. Ryan Byerly's Response to Commentaries	477
16 A Bayesian Social Platform for Inclusive and Evidence- Based Decision Making	480
S. KATE DEVITT, TAMARA R. PEARCE, ALOK KUMAR CHOWDHURY AND KERRIE MENSERSEN	
16b Commentary from Jeroen de Ridder	514
16c Commentary from Erik J. Olsson	517
16d S. Kate Devitt, Kerrie Mengersen, Tamara R. Pearce and Alok Kumar Chowdhury's response to commentaries	520
17 Measuring Social Epistemic Virtues: A Field Guide	523
MARCO MEYER	
17b Commentary from T. Ryan Byerly	543
17c Commentary from Alessandra Tanesini	546
17d Marco Meyer's Response to Commentaries	550
18 Learning from Ranters: The Effect of Information Resistance on the Epistemic Quality of Social Network Deliberation	553
MICHAEL MORREAU AND ERIK J. OLSSON	

x *Contents*

18b	Commentary from Georgi Gardiner	572
18c	Commentary from Thi Nguyen	577
18d	Michael Morreau and Erik J. Olsson's Response to Commentaries	580
19	Education as the Social Cultivation of Intellectual Virtue MICHEL CROCE AND DUNCAN PRITCHARD	583
19b	Commentary from Alessandra Tanesini	602
19c	Commentary from Lani Watson	606
19d	Michel Croce and Duncan Pritchard's Response to Commentaries	609
	<i>Index</i>	613

T&F Proofs – Not for Distribution

Figures

8.1	Summary of intellectual virtues	239
8.2	Proportion of fake news articles people find credible	241
16.1	Twitter Support tweet explaining the new feature, a prompt to encourage informed discussion. See https://twitter.com/TwitterSupport/status/1270783537667551233?s=20	484
16.2	Twentieth century horse race. Image: Casey Hibbard (25 March 2010) https://www.compelling-cases.com/how-case-studies-get-done-one-leg-at-a-time/	488
16.3	Add a hypothesis to the BetterBeliefs platform	491
16.4	Detail your hypothesis	491
16.5	Add supporting or refuting evidence	492
16.6	How the probability distribution changes for the degree of belief over the first 35 votes on a hypothesis	493
16.7	The change to probability density as votes are made on hypotheses	494
16.8	The probability density of the weight of evidence as the number of supporting and refuting evidence items of varying quality increases	494
16.9	Sample of downloadable output from the BetterBeliefs platform (authors' names withheld)	501
18.1	An open-minded, undecided and somewhat competent agent prepared to listen to a ranter who, unbeknownst to her, is on the wrong side of the debate	559
18.2	Result of a typical run of the two-person network in Figure 18.1	561
18.3	Listening to ranters on opposite sides of the issue	563
18.4	Results of a typical trial in the case with two ranters on opposite sides	564
18.5	Simulation results for the network with two ranters on opposite sides and with the ranters removed (see Figure 18.6)	564

xii *Figures*

18.6	Listening to two false ranters	565
18.7	Simulation results for the network with two false ranters (see Figure 18.6)	566
18.8	A board with four ranters, two on each side of the issue	567

T&F Proofs – Not for Distribution

Tables

8.1	Overview of Alfano et al.'s intellectual humility scale	239
8.2	Measure of conspiracist thinking	240
8.3	Regression results study 1	243
8.4	Regression results study 2	247
8.5	Regression results for individual virtues in study 2	249
8.6	Regression results individual virtues study 1	251
8.7	Demographic comparison combined sample	251
8.8	Summary statistics study 1	258
8.9	Summary statistics study 2	259
16.1	Dimensions of information quality and contributing factors for each dimension (Arazy and Kopak 2011; Mai 2013)	493
16.2	Guide to ranking evidence items on BetterBeliefs	494
16.3	Breakdown of decision quadrants: green, red, amber and white	495
18.1	Derived updating rules for belief (credence) and trust	558
18.2	Epistemic value for the board with different proportions of ranters (30 steps, average over 10,000 simulation trials)	567

T&F Proofs – Not for Distribution

Notes on Contributors

Mark Alfano is Associate Professor of Philosophy at Macquarie University.

Heather Battaly is Professor of Philosophy at the University of Connecticut.

Steven Bland is Assistant Professor of Philosophy at Huron University College.

T. Ryan Byerly is Lecturer in Philosophy of Religion at the University of Sheffield.

J. Adam Carter is Reader in Philosophy at the University of Glasgow.

Quassim Cassam is Professor of Philosophy at the University of Warwick.

Michel Croce is Assistant Professor of Philosophy at the University of Genoa.

Boudewijn de Bruin is Professor of Philosophy at the University of Groningen.

Jeroen de Ridder is Associate Professor of Philosophy at Vrije Universiteit Amsterdam and Professor by special appointment of Christian Philosophy at the University of Groningen.

Barend de Rooij is a Lecturer in the Philosophy of Humanity, Culture, and Ethics group at Tilburg University.

S. Kate Devitt is Chief Scientist of the Trusted Autonomous Systems and Associate Professor of Human-Computer Interaction at the University of Queensland.

Georgi Gardiner is Assistant Professor of Philosophy at the University of Tennessee.

Sandy Goldberg is Chester D. Tripp Professor in the Humanities and Professor of Philosophy at Northwestern University, and Professorial Fellow at the University of St. Andrews.

xvi *Notes on Contributors*

Heidi Grasswick is George Nye & Anne Walker Boardman Professor of Mental and Moral Science at Middlebury College.

Ian James Kidd is Assistant Professor of Philosophy at Nottingham University.

Colin Klein is Professor in the School of Philosophy at the Australian National University.

Alok Kumar Chowdhury is Research Associate at Queensland University of Technology.

Neil Levy is Professor of Philosophy at Macquarie University and Senior Research Fellow at the Oxford Uehiro Centre for Practical Ethics.

Kerrie Mengersen is Distinguished Professor of Statistics at Queensland University of Technology.

Marco Meyer is Junior Research Group Lead and “Freigeist” Fellow in the Department of Philosophy at the University of Hamburg.

C. Thi Nguyen is Associate Professor of Philosophy at the University of Utah.

Erik J. Olsson is Professor of Philosophy at Lund University.

Tamara R. Pearce is a PhD candidate at Queensland University of Technology.

Duncan Pritchard is Distinguished Professor of Philosophy at the University of California Irvine.

Alessandra Tanesini is Professor of Philosophy at Cardiff University.

Lani Watson is a Research Fellow in the Oxford Character Project at the University of Oxford.

Acknowledgements

Special thanks to Peter Clutton for his determined work following up with authors, copyediting contributions, and getting the final manuscript into shape. We also appreciate the patience of Andrew Beck and the staff at Routledge during what turned out to be an especially bad year for meeting deadlines.

We acknowledge the support of the Australian Research Council Grant (DP190101507, to Colin Klein and Mark Alfano) and the John Templeton Foundation (Grant #61378 to Mark Alfano). Jeroen de Ridder's work on this publication was supported by grants from the Dutch Research Council (project 276-20-024) and the Templeton World Charity Foundation.

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Dutch Research Council, The Australian Research Council, the John Templeton Foundation, or the Templeton World Charity Foundation.

T&F Proofs – Not for Distribution

Introduction

A Research Program for Social Virtue Epistemology

Mark Alfano, Colin Klein and Jeroen de Ridder

1 Why This Volume?

In recent decades, philosophers have developed a rich conceptual framework for thinking about individual epistemic virtue in general, as well as discrete epistemic virtues like open-mindedness, curiosity, intellectual humility, and intellectual courage (Turri et al. 2017). This work has sometimes been developed to help address unresolved epistemological problems and puzzles, such as the Gettier problem (Zagzebski 1996; Turri 2011) or the analysis of knowledge (Sosa 2007, 2009; Greco 2010). In addition, the field has taken on a life of its own. Even if an account of epistemic virtue does not help us formulate an account of knowledge, it is worth thinking through what it takes to be intellectually virtuous and vicious (Hookway 2006). More recently, several philosophers have developed a philosophy of epistemic vice and analyzed a range of discrete vices, such as closed-mindedness, testimonial injustice, intellectual arrogance, intellectual cowardice, and epistemic insouciance (Battaly 2014; Cassam 2016, 2018; Kidd 2016, 2018; Lynch 2018; Tanesini 2018, 2021).

In parallel, epistemology has become more social on multiple dimensions. There has been an efflorescence of research on group epistemology (can groups believe? can they know? what would it mean for them to make assertions?) (List & Pettit 2011; Gilbert 2013; Lackey 2014, 2021; Tollefsen 2015; Brady & Fricker 2016), extended knowledge (can the vehicle of knowledge extend beyond the brain and body of the knower?) (Carter et al. 2018a, 2018b), the ethics and epistemology of gossip and whistleblowing (when should one pass along testimony, and to whom?) (see, e.g., various chapters in Coady & Chase 2018), the epistemic foundations of democracy (Anderson 2006; Brennan, 2017; Landemore 2017; Goodin & Spiekermann 2018), the most fruitful structure for scientific communities and communications (Zollman 2007; Weisberg & Muldoon, 2009; Boyer-Kassem et al. 2017; O'Connor & Weatherall 2019), and a range of other topics and questions.

Meanwhile, outside of philosophy, people have begun to worry about an epistemic crisis. A 2017 lead article in the *New Scientist* proclaimed,

“Philosophers of knowledge, your time has come!”¹ In a recent interview, Barack Obama, too, suggested “We are entering into an epistemological crisis” (Goldberg 2020). As outcries about fake news demonstrate, we need a better understanding of how knowledge, ignorance, and error spread in a world characterized by communities. Conspiracy theories, too, seem to be spreading at an alarming rate, often accelerated and supercharged by technologies such as social media (Facebook, Twitter, Instagram, Reddit) and recommender systems (YouTube; Alfano et al. 2018; Klein et al. 2018). These problems are not currently well understood, but they all relate in various ways to the networked character of contemporary knowledge, error, and ignorance. How can communities and their members acquire and retain the capacities to learn from each other, including from others who disagree with them, and how can they respond rationally to conflict?

What does it take to be a good or bad epistemic agent in this contemporary environment? Within this question, we can distinguish between dispositions, behavioral patterns, and attitudes that are likely to make someone *successful* (i.e., to help them harness the wisdom of crowds, avoid the madness of masses, steer clear of fake news, broadcast their own knowledge to others in a way that secures uptake, etc.) and dispositions, behavioral patterns, and attitudes that are likely to make someone *beneficent* (i.e., to help others harness the wisdom of crowds, avoid the madness of masses, steer clear of fake news, learn from experts rather than cranks, etc.). In other words, in the context of social epistemology, there are both self-regarding and other-regarding virtues (and correlative vices). Anecdotally, we all seem to know an uncle or grandfather who tends to amplify fake news, conspiracy theories, and other epistemically problematic viral content via email, social media, and other digital interfaces. This impression was born out by a recent article, which found that there are significant individual differences in the disposition to share fake news (Guess et al. 2019). In particular, the authors found that conservative and older social media users were significantly more likely to share fake news associated with the 2016 American presidential election. Remarkably, users over the age of 65 shared *seven times* as much fake news as younger users. This demonstrates that there are meaningful individual differences in people’s social epistemic dispositions. Some are more socially epistemically virtuous (or vicious) than others.

2 The Structure of the Volume

We have divided the volume into four parts, each with a different theme: Foundational Issues, Individual Virtues, Collective Virtues, and Methods & Measurement. The divisions, and assignments of chapters to them, are impressionistic even by the standards of edited volumes. This

is as it should be. For one, a key lesson of social epistemology has been the degree to which the individual and the collective are difficult to disentangle. For another, a key lesson of virtue theory more generally has been that individual virtues often derive their value and their grounding from the interplay between individual excellence and collective flourishing (Tiberius 2018).

Further cementing this interdependence is this volume's structural experiment in collective discussion and reply. Each chapter received commentary from two other authors in the volume, and each original author then replied to these commentaries. This is a volume that attempts to practice what it preaches: the authors form part of a collective conversation about how we can know about what we know.

2.1 Foundational Issues

In the first part of the volume, we find chapters concerned with the foundational relationship between the individual and the collective, between what it takes to be a good knower and the specific capacities that might ground epistemological virtues and vices. Steven Bland's "Interactionism, Debiasing, and the Division of Epistemic Labour" suggests that different factors give rise to different sorts of epistemic vices. Broadly speaking, reliabilist vices are best handled by interventions on internal factors, while responsibilist vices should be tackled by looking at external factors.

In her "Attunement: On the Cognitive Virtues of Attention", Georgi Gardiner focuses on the cognitive role of *attention* and the unique interplay between attentional traits and epistemic virtues and vices. She argues that disproportionate attention—paid either by individuals or by groups—can be epistemically distorting even if the first-order representation of facts is impeccable. Disproportionate attention paid to, for instance, the potential downsides of an all-plant diet can distort one's overall evaluation of veganism, even if none of the particular facts or particular episodes of attention are problematic.

In "From Vice Epistemology to Critical Character Epistemology", Ian James Kidd suggests that a full picture of social virtue epistemology might move beyond individual virtues and vices to what he calls character epistemology. Focusing on epistemic vices in particular, he sketches a theory of *epistemic corruption*, on which an individual can become susceptible to patterns of epistemic vice. Corruption is initiated and stabilized by repeated interactions with bad environments and bad knowers.

Continuing the environmental theme, Neil Levy's "Narrowing the Scope of Virtue Epistemology" suggests that, while virtues and vices may be plentiful, if we focus on the *ameliorative* aims of social virtue epistemology then we may (ironically) end up focusing less on virtues and vices at all. Levy defends the idea that the environment is the best

point of intervention for many purposes and that the epistemic virtues are grounded mainly by the ways in which they bring about good epistemic environments.

Continuing a common theme, Tanesini's "Mindshaping and Intellectual Virtues" rounds out this part by focusing on the role of "mindshaping" in developing the intellectual virtues. The mindshaping literature attempts to go beyond mere theory of mind to show the ways in which predictions about others' mental states and commitments to behaving in explicable ways end up shaping our understanding of both ourselves and others (McGeer 2015). Tanesini leverages this work to argue that intellectual virtues, while in some sense individual, end up shaping and being shaped by their crucial social role.

2.2 *Individual Virtues and Vices*

Social virtue epistemology opens up the possibility that there are unexpected or unexplored virtues and vices that individuals might exhibit. There are a variety of ways in which individuals might systematically contribute to, or detract from, their epistemic environment.

In "The Vices and Virtues of Extremism", Quassim Cassam suggests that extremism is a particular way of being epistemically vicious—a *mindset*, as he puts it, that constitutes a whole pattern of "attitudes, preoccupations, emotions, and thinking patterns". Properly understood, Cassam argues, extremism is epistemically problematic independent of the background motivations, political context, or specific beliefs.

Sandy Goldberg's "Expectations of Expertise: Bootstrapping in Social Epistemology" turns to the complex role of expertise—which presumably consists in the possession of intellectual virtues, among other things—and the role of the appropriate attitudes towards expertise in the community. Otherwise-justified belief, he argues, can be undercut if there is available expertise that an agent has overlooked. By thinking of social obligations towards expertise in this way, we make possible a kind of *social-epistemic bootstrapping* in which the development of individual epistemic excellence can be translated into a better epistemic community.

Marco Meyer and Mark Alfano turn to the consequence of intellectual vice in their "Fake News, Conspiracy Theorizing, and Intellectual Vice". They present the results of a large pre-registered study showing that measures of intellectual virtue negatively correlate with belief in fake and conspiratorial news items. Notably, this is a self-assessment questionnaire. This suggests that epistemic vice, at least sometimes, need not be "stealthy".

Finally, in "Playfulness versus epistemic traps", Thi Nguyen turns to positing a specific undertheorized virtue of intellectual playfulness. He

argues that open-minded, playful examination of issues from a variety of angles is virtuous. Play is an activity done for its own sake, and is less rule-bound than comparable activities in the same space—involving a certain *lightness* with respect to the rules, as Nguyen puts it. Playfulness can be an intellectual virtue because it helps individuals avoid “epistemic traps”, a common phenomenon in which rigorous inquiry can (through no fault of an inquirer) become stuck in a small space of options.

2.3 *Collective Virtues and Vices*

While traditional virtue theory focused on individuals, there has been an increased interest in virtues and vices that can be attributed to whole groups, over and above those that are merely possessed by their individual members. Virtue epistemology is no exception, and indeed the move to *social* virtue epistemology makes it natural to focus on the epistemic virtues of communities as a whole.

In her “Solidarity: Virtue or Vice?”, Heather Battaly considers the complex, distinctively collective virtue of *solidarity*. Battaly gives several conditions that a group must have in order to count as having the trait of solidarity, including shared aims, shared goals, and a trust in the testimony of other members. Importantly, she also notes that solidarity considered as a collective *trait* might leave open whether it is virtuous or vicious in particular groups, and notes that we can find both sorts of cases.

Adam J. Carter’s “Collective (Telic) Virtue Epistemology” looks at the broader issue of collective epistemology. He draws on Ernest Sosa’s telic virtue epistemology to explicate the conditions under which a group can be said to know and argues that, contrary to what others have suggested, Sosa’s virtue epistemological framework does lend itself to an analysis of collective knowledge. There are a number of reductive and nonreductive accounts of group knowledge in the literature. Carter argues that we improve on these if we adopt a *telic* account, on which an important condition of group knowledge is that a group has the trait of committing to and aiming at knowledge.

Jeroen de Ridder’s “Three Models for Collective Intellectual Virtues” gives a synoptic review of different models of collective epistemic virtues. He notes that many models assume that there is one set of virtues that can be had both by individuals and collectives, but that it is quite possible that some virtues are distinctively collective, which is to say they can only be possessed by collectives. Byerly and Byerly’s solidarity (discussed by Battaly in this volume) is an obvious candidate, as are group traits like mutual understanding and good interpersonal deliberative practices.

Barend de Rooji and Boudewijn de Bruin continue this thread in their “Real Life Collective Epistemic Virtue and Vice”. They note that the idea of collective vice often gets real traction when we turn to praising or criticizing agents: we readily say that Boeing’s *arrogance* caused unnecessary crashes, even if no individual engineer has this vice. Thinking of collective virtues in this way also opens up interesting ameliorative possibilities, as we can begin to think about ways that organizations can scaffold collective virtues and avoid collective vices.

Finally, in her “The Social Virtue of Questioning: A Genealogical Account”, Lani Watson considers another kind of virtue that arises from the collective social practice of *questioning*. Individual questioners advance knowledge. But there is also a collective practice of asking questions and receiving answers. Done well, it can make for epistemically virtuous collectives—and indeed may be something of a foundational collective virtue.

2.4 Methods and Measurements

The final part addresses a cluster of issues that arise around the study of virtues (individual or collective) and potential interventions upon them. Social virtue epistemology has often been motivated by a strong ameliorative streak: the goal is not merely to identify the virtues that lead to good epistemic outcomes and environments but to find ways to promote and enhance them.

In “An Interdisciplinary Methodology for Studying Collective Intellectual Character Traits”, Ryan Byerly outlines a project for operationalizing epistemic traits for further study. Importantly, this assumes (as did many of the chapters in the previous section) that there are non-summativ, emergent traits of collectives. He then sketches ways in which groups might be surveyed to elicit both individual and collective attitudes, in order to discover relationships and divergences between the two.

Kate Devitt et al. focus on technological scaffolds for enhancing virtuous traits in their “A Bayesian Social Platform for Inclusive and Evidence-Based Decision Making”. They note that virtuous and vicious actions can be enhanced by technological design decisions in online platforms (echoing a theme explored in Alfano et al. (2018)), and raise the possibility of more deliberate design to promote better epistemic agents. They report on BetterBeliefs, a working proof of concept for a platform that allows for agents to pool credences in such a way that better beliefs overall can be achieved.

Marco Meyer’s “Measuring Social Epistemic Virtues: A Field Guide” offers reflections on the use of survey instruments for measuring social-epistemic virtues. He notes the need for reliable, well-validated survey instruments, especially if one is to address situationist

challenges to trait-based explanation. The chapter also explores what might need to be added to individual instruments in order to go beyond correlational evidence to causal claims about the role of the intellectual virtues.

“Learning from Ranters: The Effect of Information Resistance on the Epistemic Quality of Social Network Deliberation” provides a useful demonstration of Laputa, a powerful agent-based modeling framework for studying information flow in epistemic networks. Michael Morreau and Erik J. Olsson use this to demonstrate the counterintuitive conclusion that “ranters”—people who consistently spread misinformation—can actually benefit epistemic networks in the right circumstances. If agents can keep track of the overall reliability and anti-reliability of sources, then ranters can actually help open-minded agents calibrate on the truth.

Finally, Michel Croce and Duncan Pritchard’s “Education as the Social Cultivation of Intellectual Virtue” outlines a framework within which virtue-based models of education might promote intellectual excellence. They note the important role of intellectual exemplars—that is, of people who consistently exemplify the intellectual virtues to an above-average degree. The recognition and deployment of exemplars in an educational context is a social project and one that might play an important role in scaffolding and developing the virtues discussed in this volume.

3 A Tentative Taxonomy

The chapters in this volume span a variety of different virtues and vices. As with any field in its infancy, social virtue epistemology is still exploring its conceptual terrain. That said, we think that a new research field often benefits from a sort of rough taxonomy. We conclude by suggesting one way in which one might carve up the social-epistemic virtues (and their corresponding vices).

At a first pass, we might distinguish between two orientations had by virtues: self-regarding and other-regarding. Self-regarding virtues are those with a primary aim of enhancing one’s own epistemological position in a social-epistemic network; other-regarding virtues aim to improve the position of others in their network. As several chapters in this volume note, this is more a matter of emphasis than a hard-and-fast demarcation. Social-epistemological virtues often improve the individual in ways that help the group, and vice versa.

Crossed with each of these are three *activities* which the virtues promote: *monitoring* one’s epistemic position by keeping track of the quality of the information that flows through a social-epistemic network; *adjusting* one’s epistemic position by tweaking the trust one gives to various sources in one’s network; and *ameliorating* one’s epistemic position

by changing the structure of one's network: adding or deleting sources and adding, deleting connections between sources, or changing one's network altogether. The cross between orientations and activities gives a six-way cut on virtues. In addition, any of the virtues (we assume) can be held by both individuals and collectives. Thus, we have a 12-way potential taxonomy of social-epistemological virtues. Beyond that, each virtue presumably has at least one correlative vice, making for a 24-way taxonomy.

Some of these virtues and vices have been enumerated and described. Other cells remain blank on the map for future exploration. We thus sketch the different possibilities (conjoining individual and collective, as well as virtue and vice, for the sake of brevity).

Self-regarding monitoring: In order to benefit from the knowledge embodied in one's social network, one should understand the structure of that network. Are the people I hear from all amplifying a message from a single source, or are they independent? In the latter case, I may be able to benefit from the wisdom of crowds, as the Condorcet Jury Theorem and related proofs indicate. In the former, I may not. Knowing how my social network is structured requires ongoing vigilance—and, as Gardiner notes in her contribution, the right sort of *attention*. By contrast, neglecting to monitor the structure of my social network is liable to make me epistemically insecure. In addition, I can only benefit from the wisdom of crowds if the independent sources I listen to are sufficiently reliable. This requires keeping track of their record of verisimilitude in different domains and contexts, so that ranters' testimony can be safely disregarded as Morreau and Olsson discuss. By contrast, neglecting to monitor epistemic track records is liable to make me epistemically insecure. Watson's contribution to questioning suggests a way in which interrogative practices might similarly be seen as a form of self-regarding monitoring at the collective level.

Other-regarding monitoring: Likewise, I may be able to benefit others by recommending sources to them (or telling them to stop listening to certain sources). But I can only do this if I monitor the structure of their social networks and the epistemic track records of their sources. This is challenging, potentially privacy invading, and time consuming. It takes real effort to embody this other-regarding monitoring virtue. However, if I fail to do so, I may leave others epistemically vulnerable.

Self-regarding adjustment: Every real social epistemic network is imperfect, at least to some extent. If I manage to monitor the structure of my own network sufficiently well, I may be able to adjust my credences to account for its imperfections. The monitoring virtue is thus conceptually prior to the adjusting virtue. And the two are distinct. In principle, I could monitor adequately without being disposed to take

the imperfections I track into account when updating my beliefs. This would be a social epistemic vice. Likewise, I could monitor the epistemic track records of my sources adequately without being disposed to distrust those who have proven themselves unreliable. Again, this would be a social epistemic vice. This adjustment process might be relatively formal or might take the form of what Nguyen calls playfulness—a willingness to stay open to possible adjustments. Similarly, the collective virtue of solidarity discussed by Battaly suggests a way in which groups might perform a kind of self-regarding adjustment in response to collective concerns.

Other-regarding adjustment: Similarly, I may be able to benefit others by suggesting that they put more or less trust in various sources located in their social epistemic network. Contrariwise, I may be able to harm them epistemically by making opposite suggestions. The ability to do so depends on other-regarding monitoring dispositions, but exercising that ability (ir)responsibly is its own epistemic virtue or vice. Proper other-regarding adjustment might involve the right sort of technological scaffolding as Devitt et al. emphasize in their contribution.

Self-regarding amelioration: While all real social epistemic networks are imperfect, sometimes they are so flawed that they need to be modified. Networks can (to some extent) be *rewired*. This could involve seeking out new sources, no longer listening to sources one had previously trusted, building connections between previously unconnected sources, effecting more distal changes in the structure of the network, or, most radically, abandoning one's network altogether and plugging into another one. Doing this well depends on sufficiently successful monitoring (virtues in group 1), recognition that attempts to adjust credences are not up to the task (virtues in group 3), and the motivation and capacity to identify efficient and effective changes that one actually has the power to enact. The latter dispositions are components of ameliorating self-regarding social epistemic virtues. And, as with the other dispositions in this taxonomy, one could embody correlative vices instead of virtues. One could, for instance, be disposed to cut oneself off from reliable testifiers, plug oneself into networks that amplify fake news and conspiracy theories, and so on.

Other-regarding amelioration: Finally, just as there are self-regarding virtues and vices related to rewiring one's social epistemic network, so there are other-regarding virtues and vices related to rewiring other people's social epistemic networks. Levy's contribution suggests that many apparent failings of others are best approached as opportunities to improve a social environment. Similarly, Croce and Pritchard's emphasis on the role of intellectual exemplars might be seen as a call for the development of a corresponding series of other-regarding ameliorative virtues. Getting other people to stop trusting reliable sources and to

plug themselves into amplifiers of fake news and conspiracy theories is a practice often employed by sexual harassers and abusers, perpetrators of financial and academic fraud, and other epistemically malign actors. By contrast, being disposed to help others rewire their trust (and distrust) networks so that they are epistemically better off and less vulnerable is an other-regarding social epistemic virtue.

This taxonomy is tentative; it may not be comprehensive, and it may neglect some important distinctions. Nevertheless, the fact that many of the virtues discussed in this volume find a home there suggests that the taxonomy picks out real dispositions with significant epistemic, social, and political impact. Regardless of the ultimate taxonomy, however, we hope that this volume convinces the reader that social virtue epistemology is already a vibrant subfield, uncovering new domains and novel and interesting points of intervention on our epistemological lives.

Note

- 1 See <https://www.newscientist.com/article/mg23431194-000-philosophers-of-knowledge-your-time-has-come/>, accessed 25 August 2019.

References

- Alfano, M., Carter, J. A., & Cheong, M. (2018). Technological seduction and self-radicalization. *Journal of the American Philosophical Association*, 4(3): 298–322.
- Anderson, E. (2006). The epistemology of democracy. *Episteme*, 3(1–2): 8–22.
- Battaly, H. (2014). Varieties of epistemic vice. In J. Matheson & R. Vitz (eds.), *The Ethics of Belief*. Oxford: Oxford University Press.
- Boyer-Kassem, T., Mayo-Wilson, C., & Weisberg, M. (eds.). (2017). *Scientific Collaboration and Collective Knowledge: New Essays*. Oxford: Oxford University Press.
- Brady, M. S., & Fricker, M. (eds.). (2016). *The Epistemic Life of Groups*. Oxford: Oxford University Press.
- Brennan, J. (2017). *Against Democracy*. Princeton, NJ: Princeton University Press.
- Carter, J. A., Clark, A., Kallestrup, J., Palermos, S. O., & Pritchard, D. (eds.). (2018a). *Extended Epistemology*. Oxford: Oxford University Press.
- Carter, J. A., Clark, A., Kallestrup, J., Palermos, S. O., & Pritchard, D. (eds.). (2018b). *Socially Extended Epistemology*. Oxford: Oxford University Press.
- Cassam, Q. (2016). Vice epistemology. *The Monist*, 99(2): 159–180.
- Cassam, Q. (2018). *Vices of the Mind: From the Intellectual to the Political*. Oxford: Oxford University Press.
- Coady, D., & Chase, J. (eds.). (2018). *The Routledge Handbook of Applied Epistemology*. London: Routledge.
- Gilbert, M. (2013). *Joint Commitment: How We Make the Social World*. New York: Oxford University Press.

- Goldberg, J. (2020). Why Obama fears for our democracy. *The Atlantic*, November 16. <https://www.theatlantic.com/ideas/archive/2020/11/why-obama-fears-for-our-democracy/617087/>.
- Goodin, R. E., & Spiekermann, K. (2018). *An Epistemic Theory of Democracy*. Oxford: Oxford University Press.
- Greco, J. (2010). *Achieving Knowledge*. Cambridge: Cambridge University Press.
- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1): eaau4586.
- Hookway, C. (2006). Epistemology and inquiry: The primacy of practice. In S. Hetherington (ed.), *Epistemology Futures*. Oxford: Oxford University Press.
- Kidd, I. J. (2016). Charging others with epistemic vice. *The Monist*, 99(3): 181–197.
- Kidd, I. J. (2018). Deep epistemic vices. *Journal of Philosophical Research*, 43: 43–67.
- Klein, C., Clutton, P., & Polito, V. (2018). Topic modeling reveals distinct interests within an online conspiracy forum. *Frontiers in Psychology*, 9(189): 1–12.
- Lackey, J. (ed.) (2014). *Essays in Collective Epistemology*. New York: Oxford University Press.
- Lackey, J. (2021). *The Epistemology of Groups*. New York: Oxford University Press.
- Landemore, H. (2017). *Democratic Reason*. Princeton, NJ: Princeton University Press.
- List, C., & Pettit, P. (2011). *Group Agency*. Oxford: Oxford University Press.
- Lynch, M. (2018). Arrogance, truth and public discourse. *Episteme*, 15(3): 283–296.
- McGeer, V. (2015). Mind-making practices: The social infrastructure of self-knowing agency and responsibility. *Philosophical Explorations*, 18(2), 259–281.
- O'Connor, C., & Weatherall, J. O. (2019). *The Misinformation Age*. New Haven, CT: Yale University Press.
- Sosa, E. (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. New York: Oxford University Press.
- Sosa, E. (2009). *Reflective Knowledge: Apt Belief and Reflective Knowledge, Volume II*. New York: Oxford University Press.
- Tanesini, A. (2018). Epistemic vice and motivation. *Metaphilosophy*, 49(3): 350–367.
- Tanesini, A. (2021). *The Mismeasure of the Self: A Study in Vice Epistemology*. Oxford: Oxford University Press.
- Tiberius, V. (2018). *Well-Being as Value Fulfillment: How We Can Help Each Other to Live Well*. New York: Oxford University Press.
- Tollefsen, D. P. (2015). *Groups as Agents*. Cambridge: Polity.
- Turri, J. (2011). Manifest failure: The Gettier problem solved. *Philosophers' Imprint*, 11(8): 1–11.
- Turri, J., Alfano, M., & Greco, J. (2017). Virtue epistemology. In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/win2021/entries/epistemology-virtue/>.

12 *Mark Alfano et al.*

Weisberg, M., & Muldoon, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science*, 76(2): 225–252.

Zagzebski, L. (1996). *Virtues of the Mind*. Cambridge: Cambridge University Press.

Zollman, K. J. (2007). The communication structure of epistemic communities. *Philosophy of Science*, 74(5): 574–587.

T&F Proofs – Not for Distribution

Part I

Foundational Issues

T&F Proofs – Not for Distribution

T&F Proofs – Not for Distribution

1 Interactionism, Debiasing, and the Division of Epistemic Labour

Steven Bland

The psychological literature on cognitive biases has been a fecund source of philosophically significant controversies for the last five decades. Among the issues that divide its practitioners is the source of biased cognition. *Internalists* think that biases typically result from the operation of sub-optimal psychological processes. This camp includes psychologists working within the heuristics and biases paradigm, who blame biased cognition on our favouring efficient heuristics over sound reasoning. This paradigm fits well with virtue theoretic accounts of cognitive biases as manifestations of epistemic vices. *Externalists* claim that biases are usually the result of environmental conditions, rather than inherent features of human psychology.¹ Gigerenzer and others argue that putatively biased judgements are often artefacts of hostile informational environments. Mercier and Sperber contend that individuals perform poorly on reasoning tasks because these tasks are undertaken in isolation, rather than in dialectical engagement with others. These views suggest that virtue-theoretic treatments of cognitive bias should be contextualist and/or collectivist.

In addition to giving rise to debates about the nature of epistemic virtues and vices, this divide has spawned disagreement about how best to attenuate the vices associated with biased cognition. Internalists generally favour debiasing strategies that intervene at the level of biased minds (inside strategies), while externalists favour strategies that intervene at the level of hostile environments (outside strategies). This disagreement is the focal point of my chapter.

It seems uncontroversial at this point to say that *both* internal *and* external factors are to blame for cognitive biases. But the fact that they *interact* in complex ways, producing non-linear effects, suggests that no straightforward combination of inside and outside strategies will adequately succeed in its ameliorative purpose. For example, attempts to mitigate myside bias in individuals can blunt the debiasing power of collective deliberation. We want lawyers to be biased in favour of the positions they're defending, so they will critically vet one another's arguments more thoroughly than they would if they were impartial. What's required, then, is a *coordinated* approach that harnesses the interactions

between internal and external factors. My principal aim in this chapter is to offer one plausible plank in such an approach. My main claim is that various sources of bias are profitably handled by distinct strategies: reliabilist vices are best addressed by inside strategies, while responsibility vices are best addressed by outside strategies.² This division of cognitive labour has important consequences for institutional design and educational reform. In particular, it calls into question central claims within the growing literature on the role of education in cultivating intellectual virtues.

1 Internalism

The internalist orientation of the heuristics and biases paradigm meshes well with virtue theories that explain epistemic successes and failures in terms of the virtues and vices manifested by individual agents. In this section, I argue that the research within this paradigm indicates the need for an epistemic theory that recognizes both responsibility and reliabilist virtues/vices. According to this view, biased cognition can be ameliorated by the cultivation of these virtues in tandem, a task that virtue theorists think is best accomplished by means of proper instruction and habituation.

The heuristics and biases paradigm was generalized in the dual-process model of cognition, which distinguishes between two sources, or types, of cognition: Systems 1 and 2. System 1 produces representations automatically, involuntarily, efficiently, and in parallel. It does so with little or no cognitive strain, and without our being aware of how it does so. Most importantly, System 1 is unreflective and innumerate: it is insensitive to the quantity and quality of evidence that bears on our judgements. By contrast, System 2 processing is effortful, slow, computationally costly, and serial. It can assess evidence consciously and deliberately, albeit with more strain and cognitive resources. Thus, we represent the world with some combination of intuitions and reflective judgements; the balance between them is determined by an efficiency-accuracy trade-off. System 1 is our default mode of cognition because it is fast and efficient; System 2 has the final say on our judgements and decisions because its deliverances are generally more accurate. Cognitive biases result when System 2 fails to correct the deliverances of System 1. This is what happens when many people answer the following question (Kahneman and Frederick 2002):

A bat and ball cost \$1.10
 The bat costs one dollar more than the ball.
 How much does the ball cost?

The intuitive response is 10 cents, which is most peoples' answer. A simple calculation reveals that this answer is incorrect – the correct answer

is 5 cents – which means that most people are not sufficiently engaged in System 2 processing when answering the question. Keith Stanovich calls this propensity to over-rely on System 1 *cognitive miserliness* (Stanovich 2011).

This model suggests that biased thinking is the result of insufficient reflection: we avoid cognitive biases by taking System 1 offline and using System 2 to generate more accurate representations. To the extent that this cognitive decoupling admits of conscious control and motivational influence, its consistent practice is a prime candidate for a responsibility virtue (Samuelson and Church 2015, 1107). Samuelson and Church label the virtue of properly decoupling from the representations of System 1 *epistemic humility*; the failure to do so they classify as the vice of *epistemic arrogance*. Roberts and West (2015) similarly focus on cognitive miserliness, and advocate for the virtues of *self-vigilance* and *intellectual vitality*: to avoid biased cognition, we must know when System 1 is likely to lead us astray (self-vigilance), and engage System 2 in those conditions, to consider evidence beyond our intuitions (intellectual vitality).³

While the disposition to engage System 2 when needed is necessary to prevent or correct the cognitive biases that result from System 1 processing, Stanovich argues that it is insufficient. In addition, System 2 must have the requisite cognitive resources to make these corrections:

An aspect of dual-process theory that has been relatively neglected is that successful Type 2 override operations require both procedural and declarative knowledge. Although taking the Type 1 response priming offline might itself be procedural, the process of synthesizing an alternative response often utilizes stored knowledge of various types.

(Stanovich 2011, 95)

To give the correct answer to the bat and ball problem, we must not only stifle the intuitive answer, but generate the correct answer. Accomplishing the latter task requires that we know how to perform the necessary arithmetical calculations. Stanovich uses the term ‘mindware’ to denote the knowledge, rules, procedures, and strategies that System 2 uses when overriding the deliverances of System 1. Some biases result from the mindware used by System 2, rather than our disinclination to engage those resources; Stanovich calls these cases of *mindware problems*. When the mindware we use fails to correct the deliverances of System 1, the source of our biases is not cognitive miserliness, but a problem with our mindware. For example, if we don’t know how to compute the probability of independent events, we will likely fall prey to the gambler’s fallacy, no matter how thoroughly we scrutinize our intuitions.⁴ Though many cognitive biases result from both cognitive

miserliness and mindware problems, Stanovich insists that we recognize their distinct contributions, to better understand the distinct sources of cognitive bias, and design more effective debiasing strategies.

Stanovich's elaboration of the dual-process model requires a similar elaboration of epistemic theory. What's needed is not more responsibility vices and virtues, but the addition of an entirely different theory: virtue reliabilism. This is obvious given Stanovich's distinction between sound and contaminated mindware. Contaminated mindware includes superstitious thinking; an over-reliance on folk wisdom; and a belief in the superiority of intuition. Good mindware includes logical inference; statistical reasoning; and experimentation. Mindware is sound when its use *reliably yields accurate beliefs*, and contaminated when its use fails to do so (ibid. 193). Thus, manifesting the *ability* to use good mindware is a reliabilist virtue, and failing to do so is a reliabilist vice. But having a reliable competence is no guarantee that we will exercise it whenever we should; the bat and ball problem is a case in point: everyone can calculate the correct answer, but most people endorse the intuitive answer without doing the calculation. Avoiding cognitive bias requires that we not only possess sound mindware, but manifest the *disposition* and *motivation* to use it discriminately. Our successes in doing so seem attributable to virtuous character traits (intellectual vitality; self-vigilance; epistemic humility), and our failures seem attributable to responsibility vices (epistemic arrogance; intellectual laziness). It seems, then, that a *hybrid* virtue theory will do the best job of capturing the epistemic norms that have emerged from the heuristics and biases research on judgement under uncertainty.

Furthermore, the theory must be *holistic*, since minimizing biased cognition requires that reliabilist and responsibility virtues be manifested *together*. Cognitive decoupling is only as good as the mindware it uses, but possessing sound mindware is no help if it's not used when needed. In other words, our cognition is fragile with respect to the causes of bias: *either* reliabilist *or* responsibility vices are *sufficient* to yield systematically inaccurate judgements. This pessimistic insight may explain why cognitive biases seem so commonplace. It also highlights the importance of developing strategies that effectively cultivate both types of epistemic virtues.

Given that virtue theorists blame cognitive biases on *internal* factors, it is hardly surprising that they usually endorse what Trout calls *inside* debiasing strategies: "An *inside strategy* is a voluntary reasoning process designed to improve the accuracy of judgment by creating a fertile corrective environment *in the mind*" (Trout 2005, 418).⁵ These strategies of developing corrective virtues typically consist of some combination of instruction and habituation. With respect to the former, Roberts and West claim that we are more likely to be intellectually vital and self-vigilant when we appreciate our susceptibility to cognitive biases. They emphasize

that because most people lack this knowledge, “Our proposal depends crucially on education” (Roberts and West 2015, 2562). Courses on critical thinking and the psychology of human judgement can teach students when they’re likely to be biased, and what they can do about it. In addition, courses on formal logic, statistics, and economics, among others, can provide students with some of the mindware needed to make better judgements and decisions. Thus, a curriculum that targets the epistemic vices responsible for cognitive biases, and the epistemic virtues capable of correcting them, is often a key feature of virtue theoretic programs that seek to ameliorate the problem of cognitive bias.

However, it isn’t enough to have the *ability* to identify, avoid, and correct cognitive biases in compromising situations; we must also have the *inclination* to do so. This can be a tall order, since it often requires a substantial investment of cognitive effort and resources. But this needn’t be the case. Many theories emphasize the important role that habituation plays in the cultivation of virtues: the more often we behave virtuously, the easier it becomes to do so. There’s no reason to think that epistemic virtues are exceptional in this respect. With enough training, many cognitive tasks can be exported from System 2 to System 1; reading and basic arithmetic are obvious examples. If the cognitive processes required for debiasing can be trained to a level of automaticity, then our chances of performing them are more promising. This training requires a significant initial investment of time, energy, and resources, and is best guided by someone who has already been trained. As such, educational contexts are well suited to provide the instruction and training required to develop corrective virtues, and mitigate biasing vices.

For example, students should be instructed that they are particularly susceptible to confirmation bias and overconfidence when reasoning about matters on which they have pre-existing opinions or in which they have some personal stake. Once students appreciate this fact, they can be trained to use a number of debiasing techniques, such as consider-the-opposite: when you suspect that you are under the influence of confirmation bias and/or overconfidence, consider some of the reasons why your beliefs could be mistaken (Samuelson and Church 2015, 1105). By having students repeatedly engage in this process, across several domains, educators instil in them sound mindware and the propensity to properly use it.

2 Externalism

Externalists emphasize the role of situational factors in cognitive processing. They argue that any account of epistemic rationality that focusses predominantly on what happens inside our minds is bound to be incomplete and implausible. It is this myopic focus, they claim, that is responsible for the overly pessimistic conclusions within the heuristics

and biases literature. In particular, the jump from putatively irrational behaviours to irrational minds is made too often and without regard for the hostile environments in which those behaviours take place.

Gerd Gigerenzer has been one of externalism's most vocal advocates. He argues that many cognitive illusions can be made to disappear by restructuring the informational environments in which they occur (Gigerenzer 1991). For example, he shows that a number of the biases in our statistical reasoning can be mitigated or eliminated by presenting information in frequency formats rather than probability formats. Consider the following two ways of asking the same question:

(P) If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs?

(F) One out of 1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, out of every 1000 people who are perfectly healthy, 50 of them test positive for the disease.

Imagine that we have assembled a random sample of 1000 Americans. They were selected by a lottery. Those who conducted the lottery had no information about the health status of any of these people. How many people who test positive for the disease will actually have the disease? _____ out of _____.

The correct answer is 0.02 or 1/51. Only 18% of the students and staff surveyed at Harvard Medical School answered (P) correctly: half answered 0.95, and the average answer was 0.56 (Casscells et al. 1978). The culprit for this inaccuracy, Casscells et al. conclude, is base-rate neglect: most participants did not factor the prevalence of the disease into their calculations of the posterior probability. This seems like a straightforward case of missing or malfunctioning mindware. However, Cosmides and Tooby (1996) found that 76% of the Stanford undergraduates they asked answered (F) correctly. This is puzzling because (P) and (F) are asking the same question, and providing the same information. But they are framed differently. And because we naturally think of probabilities as relative frequencies, rather than mathematical probabilities, we find the information presented in (F) easier to compute than the information presented in (P).⁶ Consequently, the inaccurate answers to (P) are not the result of missing mindware, but of a poor fit between the mindware we use and the tasks that psychologists present us with. This is a mismatch problem that gets misdiagnosed as a disparity problem.

The problem with the heuristics and biases paradigm, from Gigerenzer's perspective, is that it is blind to these mismatch problems. Since it assumes a *single, invariant* set of reasoning norms, its practitioners interpret *any* departure from those norms as being irrational. Yet there are circumstances where we *should* diverge from those norms: frequentists should be puzzled by questions about single-event probabilities, and thus give non-Bayesian answers to (P). This being the case, Gigerenzer advocates an *ecological* conception of rationality, according to which reasoning strategies must be evaluated relative to the environments in which they're used. And he claims that "Cognitive virtue is, in my view, a relation between a mind and its environment, very much like the notion of ecological rationality" (Gigerenzer 2008, 18). We shouldn't assume that epistemic virtues are cross-situationally stable: whether or not an ability or trait is virtuous depends on the environments in which it is manifested.

Furthermore, we shouldn't assume that the biases subjects manifest when tested in isolation are endemic to human reasoning generally. Mercier and Sperber (2011; 2017) argue that the existence of these biases calls into question the intellectualist view that reason evolved to optimize the beliefs and decisions of *individuals*. In particular, the fact that our reasoning routinely operates under the influence of myside and confirmation bias makes it difficult for us to improve our beliefs and decisions. This enigma disappears, however, on their view that reason evolved to facilitate the transmission of information between human beings. The stability of communication requires that most of the information that gets transmitted and accepted is veridical; if deception were commonplace, then communication would impose too high a cost on potential communicators. Reason enables us to *argue* for or against the information that gets communicated, thus constituting a valuable tool for propagating and vetting this information. This argumentative theory nicely dispels the enigma of reason: if the reason is a tool for persuasion rather than optimization, then it should be biased in favour of our beliefs. And if argumentation serves to improve the quality of information that gets communicated, then we should expect it to be an effective antidote to biased cognition. And it is, claim Mercier and Sperber, because our biased minds are inept at identifying and correcting *our own* cognitive errors, but quite proficient at identifying and correcting the errors of *others*. Consequently, we can use dialogic environments to harness our cognitive limitations in ways that allow us to collectively overcome them. Argument is like testimony in this respect: just as testimony is an effective means of dividing the cognitive labour of acquiring information, interpersonal deliberation is an effective means of dividing the cognitive labour of reasoning about information. Since dialectically engaged groups typically have more cognitive resources at their collective disposal, and a greater inclination to use them, their reasoning tends to be

less biased than the reasoning of individuals. This is why, for example, deliberating groups are up to four times more likely to correctly complete the Wason four-card selection task than individual subjects (Moshman and Geil 1998; Mercier and Trouche 2015). Mercier and Sperber insist: "...the normal condition for the use of reasoning are social, and more specifically dialogic. Outside of this environment, there is no guarantee that reasoning acts for the benefits of the reasoned" (Mercier and Sperber 2017, 247). Since almost all of the experiments within the heuristics and biases tradition take place outside of this environment, we should be neither surprised nor distressed by their seemingly dire results: they do not reveal shortcomings of *human* rationality, but of *individual* rationality. And because its practitioners are guilty of this conflation, they tend to focus too much on what is going on inside the minds of individuals, and not enough on what is going on between them.

Externalism offers two reasons to be optimistic about the prospects of attenuating biased cognition. First, if human beings are contextually irrational, rather than constitutionally irrational, then outside debiasing strategies may prove effective: "An *outside strategy* identifies features of *the environment* whose presence can be manipulated to produce the most accurate or desirable available outcome" (Trout 2005, 420).⁷ Second, we might be able to implement these strategies at a relatively low cost: "I conjecture that changing environments can in fact be easier than changing minds" (Gigerenzer 2008, 16).⁸ Framing probabilities as relative frequencies is certainly less demanding than teaching people to reason like Bayesians. And it seems that minds are more readily opened with critical (and polite) conversations than with self-imposed strategies, such as consider-the-opposite. Thus, epistemic programs meant to ameliorate the problem of cognitive bias should provide guidance concerning the development of ecological and collectivist virtues, rather than focussing overwhelmingly on stable virtues that are attributable only to individual agents.

3 Interactionism

Almost no one can be found on the extremes of the internalist-externalist divide: there is general agreement that biased cognition has *both* internal *and* external causes.⁹ Consequently, it would seem that a *combination* of inside and outside strategies stands the best chance of mitigating cognitive biases. Some combinations are better than others, however. The best way to combine them depends on the ways in which personal (internal) and situational (external) factors are related to one another, and how they give rise to cognition.

The *conjunctive* approach is to develop inside and outside strategies independently, and implement them jointly. Accordingly, we should inculcate reliabilist and responsibilist virtues, and design more benign informational and collectivist environments, but these aims have little to

do with one another. This approach is appropriate if personal and situational factors are *independent*, such that “...the effect of some person[al] variable is the same, regardless of the situation the person is in, and the effect of the situation is the same, regardless of the kind of person in that situation” (Kihlstrom 2013, 794). Person-situation independence entails that inside and outside strategies won’t overlap or conflict: the cultivation of corrective virtues will make us less biased across situations, and the design of better environments will make us less biased across populations. If this is the case, then the conjunctive approach is the way to go: the more effective strategies we implement, the less biased we’ll be.

But this isn’t the case. Psychologists now agree that behaviour is largely the result of *interactions* between personal and situational factors. This has given rise to the doctrine of *interactionism*, according to which “... *situations are as much a function of the person as the person’s behavior is a function of the situation*” (Bowers 1973, 327). Benign framing and dialogic engagement tend to mitigate cognitive biases, but the extent to which they do so often depends on whose cognition is being de-biased. Framing statistical information in a frequency format will reduce base-rate errors, but is more likely to do so when the people presented with this information are highly numerate. Consequently, inside and outside strategies can reinforce one another in ways that cannot be recognized by the doctrine of person-situation independence. By the same token, they can also *interfere* with one another. By reducing confirmation and myside bias in individuals, we can undermine the bias-mitigating dynamics of critical dialogue (Mercier and Sperber 2011). Thus, a conjunction of debiasing strategies can sometimes lead to *worse* results than the pursuit of a single strategy, or no strategy at all.

Furthermore, the personal and situational determinants of behaviour tend to influence one another: the traits that people manifest are influenced by the environments in which they develop, and personal traits play a role in determining the situations that people put themselves in. Kihlstrom (2013) calls this the doctrine of *reciprocal determinism*. Reciprocal determinism requires that we recognize the possibility of hybrid strategies that are neither strictly inside nor outside (Bland 2020). We can improve the way people think by designing environments that foster epistemic virtues; this is an *outside-in* approach. For example, given that people are more likely to develop epistemic humility when they receive timely, unambiguous feedback about the accuracy of their judgements (Wilson et al. 2002), we can encourage epistemic humility by designing environments that regularly deliver such feedback, such as forecasting tournaments and prediction markets. And we can improve our surroundings by developing outward focussed virtues; this is an *inside-out* approach. For example, by cultivating intellectual courage, we can make people more likely to seek out critical feedback from others. Thus, reciprocal determinism opens strategic avenues that would

not exist if person-situation independence were true. Indeed, with so many strategies available to us, it would be inefficient to implement all of them: the conjunctive approach is sure to result in redundancies.

The relationships between personal and situational factors are complex, dynamic, and non-linear. Consequently, a coordinated approach to debiasing is not the sum of its personal and situational interventions. In particular, the conjunctive approach is ill advised because it leads to strategic conflict and overlap. This means that we should use debiasing strategies *selectively*, which gives rise to a *coordination problem*: how do we select which strategies to jointly implement? I will propose a partial solution to this problem in the following section.

4 A Division of Cognitive Labour

My proposal is that different types of strategies are best suited to cultivating different types of corrective virtues. While the reliabilist virtues capable of overcoming mindware problems are best cultivated by inside strategies, the responsibilist virtues capable of overcoming cognitive miserliness are best cultivated via situational scaffolding leveraged by outside, outside-in, and inside-out strategies. Since both types of virtues can be effective only when developed in tandem, all four strategies are essential to a well-coordinated approach to debiasing, though they should generally have distinct targets.¹⁰ I have two reasons for this position. First, responsibilist virtues are more difficult to cultivate in individuals than reliabilist virtues. Second, responsibilist virtues are more unstable than reliabilist virtues, in large part because the former are *emergent* features of group cognition, whereas the latter are more likely to *aggregate* in epistemic collectives. I develop these arguments below.

There is a growing empirical literature on the difficulty of debiasing that suggests that most biases cannot be personally overcome by developing responsibilist virtues, such as intellectual vitality and self-vigilance. Being intellectually vital doesn't require that we use System 2 all, or even most, of the time. Rather, we must be *selectively* vital, i.e., self-vigilant. To be properly self-vigilant, we must know when we're likely to be biased, and herein lies a problem. Most biases arise from System 1 processing, yet this processing is largely closed to introspection. The result is that most biases go undetected. Yet we don't realize that we lack internal signs of biased cognition. These two facts conspire to produce a *bias blindspot*, i.e., our tendency to more readily recognize biased thinking in others than in ourselves (Pronin et al. 2002; Pronin and Kugler 2007). And if we don't often recognize when we're biased, we're unlikely to initiate any process to remedy the situation.

Roberts and West might reply that this is one of the problems that their virtue-based education is supposed to fix: it can teach us to look

for *external* signs of biased cognition, so that we can recognize biases as readily in ourselves as we can in others. For example, we might be conditioned to consult quantitative empirical data when attempting to determine the frequency of spectacular events – mass shootings; terrorist attacks; etc. – to avoid the biases that result from relying on the availability heuristic.

Unfortunately, says Kahneman, “...this sensible procedure is least likely to be applied when it is needed most” (Kahneman 2011, 417). The problem with such inside strategies is straightforward: they require *biased* minds to do the debiasing. We are biased in favour of easy intuition over difficult deliberation, yet we’re motivated to see ourselves as rational, rigorous, and accurate thinkers (Kunda 1990). And we are biased in favour of information that confirms our positive self-image. Consequently, our standing assumption that our thinking is unbiased often survives obvious cues to the contrary. In fact, being alerted to the possibility that we’re biased can make us *more confident* in our biased judgements: it gives us another occasion to look for reasons supporting our objectivity (Hirt and Markman 1995; Sanna et al. 2002; Frantz 2006).

Yet experimental studies of particular debiasing techniques seem to provide grounds for optimism. For example, studies have found that tracking accuracy and perspective-taking can mitigate overconfidence: by keeping score of their judgemental accuracy and deploying the consider-the-opposite strategy, subjects were better able to calibrate their levels of confidence (Fischhoff 1982; 2002; Arkes et al. 1987). This is doubly good news since overconfidence leads not only to biased judgements, but biased judgements about one’s own cognitive performance. Fostering epistemic humility, then, can serve the dual purpose of reducing bias at the level of cognition, and improving our odds of identifying biased cognition at the metacognitive level.

Critics are quick to point out, however, that these interventions have been implemented in laboratory settings that do not resemble the normal conditions in which individuals formulate and think about their judgements. Ahlstrom-Vij reports that in experiments where feedback was found to reduce overconfidence,

...subjects (a) answer several questions about their degree of calibration directly after having performed the relevant judgment tasks; (b) consult graphical representations of how well their answers correspond to their actual degree of calibration; and then (c) answer several questions about what the relevant graphs suggest about their degree of overconfidence, to ensure that the subjects understand the feedback information.

(Ahlstrom-Vij 2013, 28)

The chances of ordinary people routinely seeking and receiving this type of feedback over their normal course of affairs are remote. Where feedback of this kind is available, it is typically the result of well-designed cognitive environments, such as forecasting tournaments and prediction markets. Thus, experimental subjects are not being trained to keep score of their judgements, but to integrate score-keeping information into future judgements. While this is no doubt a valuable skill for mitigating biased cognition, it is bound to remain dormant in the absence of feedback mechanisms that frequently operate *outside* a subject's control.

Kenyon and Beaulac make a similar point about the consider-the-opposite strategy. In experimental settings, subjects are *prompted* to entertain alternative perspectives, and *presented* with information that makes it easier for them to do so. They argue that this is essential to the strategy's empirical success, which casts serious doubt on its effectiveness outside of laboratory settings (Kenyon and Beaulac 2014, 347).¹¹ Once again, the problem with self-deployed strategies is that they are subject to some of the same biases that they're meant to attenuate. Consider-the-opposite needn't be a self-deployed strategy, however. In fact, perspective-taking is more readily accomplished by dialogic interaction with others. The perspectives that get entertained in such conversations are less likely to be biased because their participants aren't uniformly biased. Other people don't share our ego-centric biases because they don't share our egos: they have no stake in our objectivity. And since they typically attend to our behaviour more critically than we do, they're more likely to notice and counteract our departures from sound reasoning. This is why *collectives* can be more intellectually vital, vigilant, and humble than individuals. And this is true even when the membership of collectives doesn't manifest these virtues individually.

Consider the virtue of active open-mindedness (AOM), i.e., the tendency to thoroughly seek out and process evidence that bears on our beliefs. Actively open-minded people are less subject to myside and confirmation bias. Philip Tetlock has found that this trait is disproportionately possessed by individuals who are unusually proficient at accurately forecasting socio-political events, so-called *superforecasters*. Tetlock's Good Judgement Project tested the forecasting acumen and AOM of individuals and teams of forecasters. Unsurprisingly, high-AOM teams outperformed low-AOM teams. More surprising are the results about the makeup of high-AOM teams: Tetlock and his colleagues found that they were not necessarily made up of high-AOM members. Rather, AOM is an *emergent property* of opinionated collectives that have a common interest in the truth (Tetlock and Gardner 2015, 207–208). This is precisely what one would expect from Mercier and Sperber's collectivist perspective. Nor would they be surprised that teams outperformed individuals by a significant margin. But they go a step further, claiming that

the manifestation of virtues like AOM at the level of individuals often *interferes* with their manifestation at the level of collectives. Solutions to complex cognitive problems require a level of information collection and processing that individuals cannot readily meet. Consequently, when each team member is open to all of the relevant information, every team member is in danger of engaging that information at a superficial level. By contrast, in doxastically diverse groups whose members are subject to confirmation bias, the tasks of collecting and processing relevant information get efficiently divided: everyone devotes their limited cognitive resources to the information that best fits their existing views. As long as all of this information gets shared and critically discussed, these groups will be *more* effectively open-minded than groups with open-minded members.¹² In other words, some of the responsibilist virtues that mitigate biased cognition in individuals are not only non-summative in epistemic collectives, they're *subtractive*.¹³ So even if they could be cultivated using inside strategies, our doing so would come at the expense of more effective collectivist interventions. As we shall see, the same is not true of reliabilist virtues.

Before making this case, I should emphasize that reliable mindware seems more easily developed through inside strategies than responsibilist virtues. Richard Nisbett and his colleagues conducted a series of studies that suggest that peoples' reasoning abilities can be improved by teaching them formal rules of inference. In a longitudinal study that tested undergraduates' statistical-methodological and conditional reasoning in their first and fourth years of study, Lehman and Nisbett (1990) found that students studying psychology and social science experienced a much greater improvement in their statistical-methodological reasoning than students studying natural science and humanities, while conditional reasoning improved much more in the latter groups than in the former. The first disparity can be explained by the disproportionate training that psychology and social science students receive in statistical reasoning in uncertain domains. Lehman and Nisbett conjecture that the improvement in the conditional reasoning of natural science students is due to their training in mathematics, though they remained puzzled about a similar improvement in humanities students. Lehman et al. (1988) found a similar pattern in the effects of graduate instruction. Their cross-sectional study compared the performance of first-year and third-year students in chemistry, law, medicine, and psychology on a questionnaire that required them to use statistical-methodological and conditional reasoning to solve a variety of scientific and everyday problems. They found a significant improvement from first to third year in the performances of medicine and psychology students, but not in those of chemistry and law students.¹⁴ They obtained the same result in a longitudinal study that compared the performances of students in the first and third years of their programs.

Not everyone is as optimistic about the prospects of mitigating cognitive biases by means of statistical instruction. Tversky and Kahneman note that,

Misconceptions of chance are not limited to naïve subjects. A study of the statistical intuitions of experienced research psychologists revealed a lingering belief in what may be called the “law of small numbers,” according to which even small samples are highly representative of the populations from which they are drawn.

(Tversky and Kahneman 1974, 1125–1126)

Learning to reason in accord with the law of large numbers, it seems, does not eliminate the tendency to use the representativeness heuristic.

This fact alone does not impugn the efficacy of formal training for two reasons. First, it is compatible with the possibility that well-trained individuals are *less* susceptible to statistical biases than untrained individuals. The work of Nisbett et al. seems to indicate that this is the case. Second, an expert’s misuse of the representativeness heuristic is a different kind of epistemic failure than a layperson’s ignorance of the law of large numbers: the former shortcoming stems from cognitive miserliness, rather than a mindware gap. As Stanovich explains, there is an inverse relationship between mindware gaps and cognitive miserliness:

One interesting implication of the relation between miserly processing and mindware gaps is that the fewer gaps one has, the more likely that an error may be attributable to miserly processing. In contrast, errors made by someone with little relevant mindware installed are less likely to result from miserly processing than to mindware gaps.

(Stanovich 2011, 102)

As the name suggests, mindware is a cognitive tool that can confer epistemic benefits only when it is used properly. But failing to *use* it is not the same thing as failing to *possess* it. The second failure entails the first, but not *vice versa*: you can fail to use mindware that you do have, but you can’t use mindware that you don’t have. Formal training can mitigate the second failing without mitigating the first: stocking System 2 with sound reasoning techniques does not ensure that System 2 will routinely engage them when necessary. Having the capacity to reason soundly is a necessary condition for effective debiasing, but not a sufficient condition; we must also *exercise* that capacity. As I’ve already noted, reliabilist and responsibilist virtues must be manifested together to attenuate cognitive biases.

Reliabilist virtues have another feature that bolsters the recommendation that they be cultivated using inside strategies. Unlike

responsibilist virtues, they are *robustly enhancive* when manifested by individuals: they tend to promote, rather than prohibit, effective debiasing across a range of diverse environments.¹⁵ As discussed above, the primary reason why responsibilist virtues are cross-situationally unstable is that they tend not to be summative in collectivist contexts. Reliabilist virtues, on the other hand, are more likely to *aggregate* in epistemic collectives than emerge *ex nihilo*: highly numerate, inferentially savvy groups are typically made up of highly numerate, inferentially savvy members. This is supported by the finding that groups tend to solve problems that admit of demonstrably correct solutions when any of their members do so (Davis 1973; Laughlin and Ellis 1986; Bonner et al. 2002). The more reliable mindware there is within a group, the greater its chances of solving a variety of such questions without falling prey to logical fallacies or statistical biases. This may also explain Tetlock's finding that grouping superforecasters yields a greater improvement than grouping normal forecasters (Tetlock and Gardner 2015, 205): their collective mindware gets consolidated, even if some responsibilist virtues, such as active open-mindedness, do not. Thus, reliabilist virtues are more often a *precondition* than a result of productive discourse. Groups whose members collectively possess a wide range of reliable mindware are in a better position to mitigate the biases that emerge in their critical discussions than groups that lack such mindware. They will effectively do so, however, only if they manifest the responsibilist virtues required to make proper use of their cognitive resources: they must be open to multiple perspectives, modify their views in light of new information, etc. Whether or not groups manifest these virtues depends more on their makeup, motivation, and the settings of their deliberations than it does on the characteristics of their members. All of this suggests that inside strategies best equip us with relevant mindware, while situational interventions best compel us to use them appropriately.

If correct, this view has significant implications for institutional design. Institutions that wish to limit biased cognition should begin by designing cognitive environments that *harness* reliabilist virtues and *induce* responsibilist virtues. Several of our most important institutions do this well. The principal actors in legal systems require formal training in contractual and ethical reasoning before they can take part in legal deliberations. Not all actors are expected to be impartial in their use of these skills, however. Quite the reverse: each side in a legal proceeding has an advocate who is supposed to be biased in its favour. The norms that govern these proceedings ensure that each advocate has an equal opportunity to present their case, which is subject to the same procedural rules determining what constitutes acceptable evidence, arguments, objections, etc. In this way, legal proceedings divide the epistemic labour that's required to effectively implement the consider-the-opposite

strategy. The same is true of institutional science. Scientists must be extensively trained to use the mindware that's appropriate to their fields of study: statistical analysis, experimental design, empirical measurement, etc. Yet, as Popper insists, to expect scientists to engage these resources without bias is to misunderstand the source of scientific objectivity:

...what we call 'scientific objectivity' is not a product of the individual scientist's impartiality, but a product of the social or public character of scientific method; and the individual scientist's impartiality is, so far as it exists, not the source but rather the result of this socially or institutionally organized objectivity of science.

(Popper 2002[1996], 426)

Popper puts little stock in the justifications that scientists give for their own theories since these justifications are inevitably contaminated by personal biases. However, scientists are expected and incentivized to *publicize* their findings and the methods they use to arrive at them. This gives the community of scientists, who do not share the same biases, the opportunity to *criticize* one another's work within a *common framework* of rules and standards. Since scientists are bound to critically vet one another's research more thoroughly than they vet their own, this division of cognitive labour gives rise to more objective results than the results that any scientist can achieve on their own.¹⁶

This view also has important implications for educational policy. Several epistemologists have recently turned their attention to this important topic, but their general approach, I argue, is not well suited to meeting the goal of nurturing less-biased inquirers.

5 Implications for Education

Much of the contemporary work in this area focusses on the role that education should play in cultivating *responsibilist* virtues in students. This trend is welcome and understandable in light of the disproportionate attention that reliabilist virtues have received in Western pedagogical traditions. However, the strategies for promoting responsibilist virtues in educational contexts are lopsidedly *internalist* and *individualistic*. They include: explicit instruction on the virtues and vices; routinely practicing virtuous behaviours in the classroom; drawing attention to exemplars; modelling the virtues; and assigning projects that encourage reflection on and assessment of personal epistemic behaviours (Baehr 2013; Battaly 2016b; Roberts 2016). While these strategies may bear some valuable fruit in general, I am dubious of their prospects for helping students mitigate biased cognition, for the reasons articulated above. I doubt that intellectual vitality, self-vigilance, and epistemic humility can be developed to any significant degree through explicit instruction,

personal practice, the examination of exemplars, in-class modelling, and self-reflections and assessments alone. In the remainder of this section, I outline a more promising pedagogical approach to cultivating bias-mitigating responsibilist virtues.

This approach involves teaching strategies of *cognitive outsourcing*, i.e., strategies for creating and exploiting positive epistemic environments. Educators already teach students how to outsource the tasks of acquiring and vetting information by showing them how to search for testimonial knowledge (online or at a library) and establish expertise (by examining credentials, looking for peer-reviewed sources, etc.). But they don't often teach students how to outsource the cognitive abilities and/or dispositions required to productively process the information at their disposal. To this end, I have a few suggestions.

First, students should be taught to appreciate both the difficulties of personal debiasing and the positive effects of situational interventions. As we have seen, there are good reasons to be pessimistic about approaches that seek to cultivate responsibilist virtues by teaching students how to implement inside debiasing strategies on their own. In the absence of any instruction on the limits of these strategies, students are in danger of *reinforcing* their biases when employing them: they can end up more confident in the accuracy of their judgements because they believe that they've effectively debiased the cognitive processes that generated them. At the very least, inside strategies should be supplemented with lessons that identify and explain the main obstacles to personal debiasing, such as self-deception and bias blindspot. This may encourage more cautious approaches to personal debiasing, and a greater openness to environmental interventions. The effectiveness of situational interventions can be illustrated by having students complete cognitive tasks on their own, and with environmental support, so that they can actually see the differences in outcomes. For example, educators may have their students complete the four-card selection task on their own, and then in small groups. The typical increase in the proportion of correct answers from the first condition to the second can serve to dramatically demonstrate the power of collective deliberation, as well as the possible pitfalls of inside strategies of promoting intellectual vitality, self-vigilance, and/or epistemic humility.

Second, students should receive instruction on what makes environmental interventions effective at ameliorating cognitive biases. For example, students may be asked to record the reasoning that led to their answers in the four-card selection task, and then explain why they were better able to correctly complete the task when grouped with their peers. What differences in reasoning in these two conditions led to the differential success rate? This is one way of getting students to recognize the potential of collective deliberation to aggregate information, pool reliable mindware, and entertain multiple perspectives. They may then

be asked to identify other collectives that have these positive features, so that they become better able to recognize productive sources of collective deliberation.

Third, students should be taught how to *improve* cognitive environments. This will involve explicitly identifying situational features that interfere with the mitigation of cognitive biases. For example, students should be warned about the epistemic dangers of collective deliberation, including groupthink, polarization, and overconfidence. These pitfalls have many causes that can be explored through lessons and assignments: cascades; a lack of viewpoint diversity; poor incentive structures; a lack of timely, precise feedback; etc. Rather than focusing exclusively on how to mitigate biases in their own thinking, students should also be introduced to interventions that have proven successful at mitigating the biases that routinely afflict group cognition. Indeed, Kahneman laments the lack of training that individuals receive in optimizing organizational reasoning: “One example out of many is the remarkable absence of systematic training for the essential skill of conducting efficient meetings” (Kahneman 2011, 418). To this end, students could be trained to use the Delphi method of aggregating viewpoints as a way of avoiding cascades; red teaming and adversarial collaboration as ways of ensuring open-mindedness and viewpoint diversity; and the premortem as a way of priming critical thinking and constraining confidence.¹⁷ In addition, they could be taught that effective leaders are typically inquiring and self-silencing, and that successful organizations tend to value diachronic improvement over occasional success. Having learned these lessons, students might be asked to apply them by suggesting ways in which negative epistemic cultures, such as social media platforms, could be improved. Moreover, a larger proportion of course work should be done in groups, and graded not only on its outcomes, but on the *processes* that groups self-consciously implement.

Fourth, students should be introduced to positive epistemic cultures in which they can participate on a regular basis. Among these positive cultures are deliberative communities that incentivize fair, rigorous, open-minded dialogue, such as debate clubs, and online platforms like *Kialo* (<https://www.kialo.com/>) and *Change My View* (<https://www.reddit.com/r/changemyview/>). Other good examples are score-keeping cultures, such as forecasting tournaments (e.g., *Metaculus*: www.metaculus.com) and prediction markets (e.g., *PredictIt*: www.predictit.org), that promote intellectual humility, vitality, and self-vigilance. Cultures with these features can be modelled in the classroom. For example, a poker tournament in which students must evaluate their decisions retrospectively can teach them about the importance of making fine-grained probabilistic judgements, belief updating, and decision-making under uncertainty.¹⁸ This is just one of many possible outside-in debiasing strategies.

Finally, students should be encouraged to develop *outward-focussed virtues*, i.e., dispositions to seek out, create, and/or harness positive epistemic environments. One candidate is intellectual gregariousness, which Brogaard characterizes as "...a natural or automatic tendency to engage with intellectual peers for the sake of getting to the truth" (Brogaard 2019, 451). Those who possess this virtue enjoy the back-and-forth of interpersonal deliberation, and consequently are more likely to have their biases checked by interactive argumentation. Educators can foster intellectual gregariousness in their students by creating stimulating dialogical environments. This can be done by holding competitive events, such as group problem-solving tournaments and in-class debates, as occasions for students to implement the mindware that they've learned. Participation in these events should not be graded, so that students learn to hold one another accountable, rather than relying on extrinsic motivation. Ideally, students will thereby come to appreciate and enjoy engaging their peers in intellectual exchanges as a source of knowledge and cognitive self-improvement, thus employing an inside-out debiasing strategy.

This virtue-based educational approach to ameliorating the problem of cognitive bias requires that we re-think the character of epistemic virtues. It is often thought that *responsibilist* virtues are more likely to be cultivated through effortful learning and habituation than *reliabilist* virtues, which is why we are *personally responsible* for the former, but not necessarily the latter. This may also be part of epistemologists' rationale for focusing on the role of education in developing *responsibilist* virtues, rather than *reliabilist* virtues. Yet, when it comes to the virtues that facilitate the avoidance/minimization/correction of cognitive biases, this supposition gets things backwards. *Reliabilist* virtues, in the form of sound mindware, are more readily imparted to individuals by means of instruction and habituation: we can learn how to reason in ways that avoid sources of systematic error, and can be held responsible for an inability to do so. On the other hand, the tendency to use this mindware when and as appropriate cannot be as easily trained, since it depends to a greater extent on features of the *environments* in which we reason. It is difficult, therefore, to hold agents *directly* responsible for failing to manifest the virtues that are constitutive of this tendency. Nevertheless, we can and should hold individuals *indirectly* responsible insofar as they can knowingly exert control over their cognitive environments.¹⁹ Teaching students how to do this, by imparting strategies of cognitive outsourcing, should be one of the central aims of any educational approach to ameliorating the problem of cognitive bias.

Acknowledgements

I am grateful to Mark Alfano, Jeroen de Ridder, and Jon Marsh for helpful comments on an earlier draft of this chapter.

Notes

- 1 The internalism/externalism nomenclature comes from the exchange between Matheson (2006) and Gigerenzer (2008).
- 2 I am using the terms ‘reliabilist virtue/vice’ and ‘responsibilist virtue/vice’ in the inclusive way that Battaly does in (Battaly 2016a). Reliabilist virtues/vices are cognitive faculties that need not be acquired or personal, and for which we need not be responsible, but whose epistemic standing and value is determined by their truth-conduciveness. Responsibilist virtues/vices are personal character traits that can be acquired, and for which we are responsible, and whose epistemic standing and value depends at least in part on their motivational elements. I am also open to the possibility that many cognitive biases are the result of vicious thinking styles, rather than cognitive faculties or traits – on this point, see Cassam (2019, Ch. 3).
- 3 Roberts and West identify four epistemic vices responsible for cognitive biases – intellectual laziness; blinkered vision; associative coherence; and substitution – but each of them are aspects of cognitive miserliness.
- 4 Stanovich classifies this as a case of *mindware gap*, i.e., our lacking the mindware that’s necessary to correct an intuitive response. The other problem is *contaminated mindware*, which routinely overrides intuitive responses with inaccurate judgements.
- 5 However, one can be an internalist about the sources of cognitive bias, but a pessimist about the prospects of inside strategies. Kahneman seems to fit this description – see Kahneman (2011, 417).
- 6 Internalists also recognize the important role that framing effects can play in our judgements, but they draw different conclusions about the psychological processes responsible for the effects. See, for example, the classic exchange between Gigerenzer and Kahneman and Tversky in Gigerenzer (1996) and Kahneman and Tversky (1996).
- 7 The distinction between inside and outside strategies may have to be sharpened in light of the extended cognition thesis (Clark and Chalmers 1998). To this end, Alfano and Skorborg (2018) helpfully articulate a distinction between embedded, scaffolded, and extended cognition, drawing on Palermo’s (2014) view that extended cognition necessarily involves stable feedback loops between cognitive agents and their environments.
- 8 See also Trout (2005) and Ahlstrom-Vij (2013).
- 9 See, for example, Kahneman and Tversky (1996), Samuels et al. (2004), and Samuelson and Church (2015).
- 10 To be clear: this is meant to be a heuristic, rather than a categorical rule. There are some outside interventions that promote reliabilist virtues, such as Gigerenzer’s framing effects, and some inside interventions that promote responsibilist virtues. My claim is that *on balance* this is more rarely the case.
- 11 See also Trout (2005, 419–420).
- 12 This conclusion is reinforced by Zollman’s (2010) work with network models.
- 13 The epistemic vices that give rise to miserly processing are instances of what Smart (2018) calls *Mandevillian intelligence*: traits that are vicious in most solitary circumstances, and virtuous in some collectivist contexts.
- 14 The improvement in the performance of psychology students was more than double the gains made by medical students. Lehman et al. replicated their results concerning psychology and chemistry students in a cross-sectional study at an alternative institution.

- 15 To be more precise: reliable mindware tends to be epistemically enhansive across its *domain of application*. The ability to do frequentist statistics is of little use when it comes to estimating the probabilities of unique events. But this inability does not have deleterious effects on any reasoning that can address such problems. Responsibilist traits can have deleterious effects when manifested by individuals, even in the domains in which they apply.
- 16 For a similar view of scientific objectivity, see Longino (1990). I should clarify that Popper and Longino offer normative accounts of how science *should* work. There are good reasons to think that science often fails to fit their descriptions. Chief among them is the so-called ‘replication crisis’ in the behavioural and life sciences. However, the crisis has precipitated a ‘credibility revolution’ (Vazire 2018) whose inside strategies generally focus on promoting reliabilist virtues, and whose outside (inside-out; outside-in) strategies generally focus on promoting responsibilist virtues. I make this argument in greater detail in (Bland 2020).
- 17 The Delphi method requires group members to submit anonymous judgements, in the form of probability estimates, in a series of rounds, between which members can freely deliberate, until a consensus is achieved. Red teaming is the practice of creating a group whose purpose is to challenge the collective’s prevailing positions. A premortem is the exercise of having a group imagine that it has failed to meet its objective, and listing the possible explanations for the imagined failure.
- 18 For a compelling account of the role that poker can play in cultivating bias-mitigating virtues, see Duke (2018).
- 19 According to this view, praise and blame for epistemic behaviours must often extend beyond individual agents, to the parties who are responsible for the relevant features of the cognitive environments in which their behaviours take place.

References

- Ahlstrom-Vij, K. (2013). *Epistemic Paternalism: A Defense*. New York: Palgrave Macmillan.
- Alfano, M. & Skorburg, J.A. (2018). Extended knowledge, the recognition heuristic, and epistemic injustice. In D. Pritchard, J. Kallestrup, O. Palermos & J.A. Carter (eds.) *Extended Epistemology* (pp. 239–256). Oxford: Oxford University Press.
- Arkes, H.R., Christensen, C., Lai, C. & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes* 39: 133–144.
- Baehr, J. (2013). Educating for intellectual virtues: From theory to practice. *Journal of Philosophy of Education* 47(2): 248–262.
- Battaly, H. (2016a). Epistemic virtue and vice: Reliabilism, responsibilism, and personalism. In C. Mi, M. Slote & E. Sosa (eds.) *Moral and Intellectual Virtues in Western and Chinese Philosophy* (pp. 99–120). New York: Routledge.
- Battaly, H. (2016b). Responsibilist virtues in reliabilist classrooms. In J. Baehr (ed.) *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology* (pp. 163–183). New York: Routledge.
- Bland, S. (2020). An interactionist approach to cognitive debiasing. *Episteme* 19(1): 66–88. <http://dx.doi.org/10.1017/epi.2020.9>.

- Bonner, S.E., Baumann, M.R. & Dalal, R.S. (2002). The effects of member expertise on group decision making and performance. *Organizational Behavior and Human Decision Processes* 88: 719–736.
- Bowers, K.M.S. (1973). Situationsim in psychology: Analysis and a critique. *Psychological Review* 80: 307–336.
- Brogaard, B. (2019). Dual-process theory and intellectual virtue: A role for self-confidence. In H. Battaly (ed.) *The Routledge Handbook of Virtue Epistemology* (pp. 446–461). New York: Routledge.
- Cassam, Q. (2019). *Vices of the Mind: From the Intellectual to the Political*. Oxford: Oxford University Press.
- Casscells, W., Schoenberger, A. & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine* 299: 999–1001.
- Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis* 58(1): 7–19.
- Cosmides, L. & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature of judgment under uncertainty. *Cognition* 58: 1–73.
- Davis, J.H. (1973). Group decisions and social interactions: A theory of social decision schemes. *Psychological Review* 80(2): 97–125.
- Duke, A. (2018). *Thinking in Bets: Making Smarter Decisions When You Don't Have All the Facts*. New York: Penguin.
- Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic & A. Tversky (eds.) *Judgement Under Uncertainty: Heuristics and Biases* (pp. 422–444). Cambridge: Cambridge University Press.
- Fischhoff, B. (2002). Heuristics and biases in application. In T. Gilovich, D. Griffin & D. Kahneman (eds.) *Heuristics and Biases: The Psychology of Intuitive Judgement* (pp. 730–748). Cambridge: Cambridge University Press.
- Frantz, C. (2006). I AM being fair: The bias blind spot as a stumbling block to seeing both sides. *Basic and Applied Social Psychology* 28(2): 157–167.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear. *European Review of Social Psychology* 2: 83–115.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review* 103(3): 592–596.
- Gigerenzer, G. (2008). *Rationality for Mortals: How People Cope with Uncertainty*. Oxford: Oxford University Press.
- Hirt, E.R. & Markman, K.D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgements. *Journal of Personality and Social Psychology* 69: 1069–1086.
- Kahneman, D. (2011). *Thinking Fast and Slow*. London: Penguin Books.
- Kahneman, D. & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (eds.) *Heuristics and Biases: The Psychology of Intuitive Judgement* (pp. 49–81). Cambridge: Cambridge University Press.
- Kahneman, D. & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review* 103: 582–591.
- Kenyon, T. & Beaulac, G. (2014). Critical thinking education and debiasing. *Informal Logic* 34(4): 341–363.
- Kihlstrom, J. (2013). The person-situation interaction. In D. Carlston (ed.) *The Oxford Handbook of Social Cognition* (pp. 786–806). Oxford: Oxford University Press.

- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin* 108(3): 480–498.
- Laughlin, P.R. & Ellis, A.L. (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology* 22(3): 177–189.
- Lehman, D., Lempert, R.O., & Nisbett, R.E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist* 43(6): 431–442.
- Lehman, D.R. & Nisbett, R. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology* 26(6): 952–960.
- Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.
- Matheson, D. (2006). Bounded rationality, epistemic externalism, and the Enlightenment picture of cognitive virtue. In R. Stainton (ed.) *Contemporary Debates in Cognitive Science* (pp. 134–144). Oxford: Blackwell.
- Mercier, H. & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences* 34(2): 57–74.
- Mercier, H. & Sperber, D. (2017). *The Enigma of Reason*. Cambridge, MA: Harvard University Press.
- Mercier, H. & Trouche, E. (2015). Experts and laymen grossly underestimate the benefits of argumentation for reasoning. *Thinking and Reasoning* 21(3): 341–355.
- Moshman, D. & Geil, M. (1998). Collaborating reasoning: Evidence for collective rationality. *Thinking and Reasoning* 4(3): 231–248.
- Palermos, S.O. (2014). Loops, constitution, and cognitive extension. *Cognitive Systems Research* 27: 25–41.
- Popper, K. (2002 [1996]). *The Open Society and Its Enemies*. Fifth Edition. New York: Routledge.
- Pronin, E. & Kugler, E. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology* 43(4): 565–578.
- Pronin, E., Lin, D. & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin* 28: 369–381.
- Roberts, R.C. (2016). Learning intellectual humility. In J. Baehr (ed.) *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology* (pp. 184–201). New York: Routledge.
- Roberts, R.C. & West, R. (2015). Natural epistemic defects and corrective virtues. *Synthese* 192: 2557–2576.
- Samuels, R., Stich, S. & Bishop, M. (2004). Ending the rationality wars: How to make disputes about human rationality disappear. In R. Elio (ed.) *Common Sense, Reasoning, and Rationality* (pp. 236–268). New York: Oxford University Press.
- Samuelson, P.L. & Church, I.M. (2015). When cognition turns vicious: Heuristics and biases in light of virtue epistemology. *Philosophical Psychology* 28(8): 1095–1113.
- Sanna, L., Stocker, S. & Schwarz, N. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(3): 497–502.
- Smart, P.R. (2018). Mandevillian intelligence. *Synthese* 195: 4169–4200.

- Stanovich, K.E. (2011). *Rationality and the Reflective Mind*. Oxford: Oxford University Press.
- Tetlock, P. & Gardner, D. (2015). *Superforecasting: The Art and Science of Prediction*. Toronto: Signal.
- Trout, J.D. (2005). Paternalism and cognitive bias. *Law and Philosophy* 24: 393–434.
- Tversky, A. & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science* 185(4157): 1124–1131.
- Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science* 13(4): 411–417.
- Wilson, T.D., Centerbar, D.B. & Brekke, N. (2002). Mental contamination and the debiasing problem. In T. Gilovich, D. Griffin & D. Kahneman (eds.) *Heuristics and Biases: The Psychology of Intuitive Judgement* (pp. 185–200). Cambridge: Cambridge University Press.
- Zollman, K.J.S. (2010). The epistemic benefit of transient diversity. *Erkenntnis* 72: 17–35.

T&F Proofs – Not for Distribution

1b Commentary from Neil Levy

Steven Bland's suggestion that responsibilist virtues and reliabilist virtues can best be inculcated in different kinds of ways is a fascinating one. He identifies reliabilist virtues (or, more plausibly, a subset of such virtues) with "mindware", and argues that they can be inculcated via *inside* strategies, where an inside strategy is one that focuses on the mind of the individual agent. Responsibilist virtues, on the other hand (the panoply of virtues on which the majority of virtue epistemologists have focused), are best inculcated via outside strategies, for instance by structuring institutions so that our dispositions are harnessed to veritistic ends. There is a great deal to chew over in this suggestion; pursuing it further opens onto a variety of important issues and promises to be very fruitful. Here, I want to point out one problem with the proposal that might make it unpalatable to some epistemologists, and suggest a perspective from which the problem might dissipate.

Bland's proposals are developed in the service of debiasing. The biases he's concerned with are not prejudices, but predictable and (apparently) species-typical dispositions that see us (again, apparently) often departing from the canons of rationality. There's more than a hint in his chapter that he thinks of reliabilist and responsibilist virtues as debiasing agents in quite different senses. Reliabilist virtues give us the mindware to engage in logical reasoning: they enable us to be rational in *that* sense. Responsibilist virtues aim at *ecological rationality*. Mapping reliabilist virtue onto direct rationality and responsibilist virtue onto ecological rationality makes sense insofar as Bland is correct that reliabilist virtue depends more on inside strategies.

Logical reasoning is a property of our cognitive processes, and while such reasoning need not be entirely internal to agents, it is manifested in the transitions between cognitive states themselves. Only if such transitions have appropriate properties is reasoning logical in this sense. Ecological reasoning, on the other hand, does not require that the transitions between states have any particular properties at all: it requires instead that our information processing is well suited to our task, not that the processing has any particular properties. Ecological rationality

depends on a relation between processing and outcome, not the properties of the process itself.

Virtue epistemologists in the responsibilist tradition will, I predict, be reluctant to identify the virtues they prize with ecological rationality *rather than* a property of the process itself. While we want our reasoning to be successful, we also want it to owe its success to how well we've responded to evidence, not to chance or to the ways in which our environment has been structured by others. I suggest that it's realistic to aim at something more satisfying than mere ecological rationality, compatible with Bland's recognition that virtues must often be outsourced and pursued via outside strategies.

We can aim at something more than mere ecological reasoning if I'm right in denying that many of the parade ground examples of irrational processing really involve irrationality on the part of the agent. Bland opens with the "bat and ball" item from Frederick's original three-item cognitive reflection test. Most people do badly on the CRT. Since the question is one on which there is an objectively correct response and the arithmetic is quite trivial, this seems to be a good example of irrationality. Standardly, wrong answers on the CRT are said to be explained by a disposition to rely on intuition, which generates a misleading answer, rather than on effortful cognition. While something like that story may be true, we should resist the easy identification of a reliance on intuition with irrationality. CRT items are trick questions (indeed, newer CRTs have sometimes been constructed by googling "trick questions"; Thomson and Oppenheimer 2016). Trick questions work by implicating a certain response; in effect, they offer implicit testimony that that response is correct. Is it really less rational to be guided by the testimony they offer rather than to reject it and perform the arithmetic?

As Bland recognizes, a disposition to inhibit the intuitive answer isn't sufficient to generate the right answer in any case; that entails that rejecting the (apparently) recommended response is risky. We might do better to accept the testimony rather than take the risk: that might be the rational strategy. Factor in the fact that these tests are usually performed under conditions in which spending longer on an item is irrational (since there are opportunity costs) and it might well be those who actually take the time to do the arithmetic who should be seen as irrational. Set that issue aside: the important thing to see how is that in being guided by testimony, the person who gets the CRT wrong is responding to *evidence*. Testimony is evidence, and it's rational to alter our credences in its light. The CRT is a case in which we have conflicting evidence, and it's not obvious who the rational agent is: the one who accepts the testimony or the one who rejects it and goes on to do the calculation (of course, *having* done the calculation the person who rejected the testimony can see that the testimony is misleading, but it's far from obvious that it is rational to probe testimony in this kind of way, in low stakes situations like this one).

Other examples are even clearer. Take framing effects. Framing works to implicitly recommend options (Fisher 2020; Levy 2019). Ordinary agents make options salient by framing when they think highly of them, and those who alter their preferences in response to framing do so in a way that reflects the actual evidential force of the testimony thereby given. There's nothing irrational about this: far from it. Again, we *want* people to alter their credences in response to testimony. Similar stories can be told about a range of alleged biases: they work through responses to implicit testimony. The prestige bias and conformity bias can both be understood along these lines, for example (Levy 2022).

This perspective offers a different way of thinking about outsourcing, about structuring epistemic environments and the other kinds of outside strategies Bland recommends. We need not see them as aiming at ecological rationality. We should structure environments and outsource cognition so that reliable evidence is offered to agents. That's not (merely) a way of increasing the likelihood of them getting things right; it's a way of making them more likely to get things right *by* responding to the actual value of the overall evidence. There's no conflict between inside and out; not when the environment is appropriately structured. It is only in epistemically hostile environments that such conflicts arise.

References

- Fisher, S.A. 2020. Rationalising framing effects: At least one task for empirically informed philosophy. *Crítica: Revista Hispanoamericana de Filosofía* 52: 3–30.
- Levy, N. 2019. Nudge, nudge, wink, wink: Nudging is giving reasons. *Ergo: An Open Access Journal of Philosophy* 6. <https://doi.org/10.3998/ergo.12405314.0006.010>.
- Levy, N. 2022. *Bad Beliefs: Why They Happen to Good People*. Oxford University Press.
- Thomson, K.S. & Oppenheimer, D.M. 2016. Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making* 11: 99–113.

1c Commentary from Michel Croce and Duncan Pritchard

Virtue Responsibilism, Mindware, and Education

Understanding and counteracting the negative effects of biased cognition currently represents a major challenge for psychologists and philosophers interested in how human beings think. In his chapter, Steven Bland sheds light on the complexity of the challenge and offers an insightful ameliorative approach to handling the problem of cognitive biases from a virtue-theoretic perspective, concluding with a focus on the educational strategies that can help students acknowledge and counter the effects of biased cognition.

Biased cognition is an obvious source of epistemic vice, but there is some controversy about whether cognitive biases generate reliabilist or responsibilist epistemic vices. Bland's argument, in a nutshell, is that since the development of cognitive biases is due to the interplay of internal psychological processes and external (i.e., environmental) conditions, it cannot be expected that a solution to the problem tackles only one of these dimensions. According to Bland, the most promising way to counteract our proneness to biased cognition involves a *coordinated* approach that divides the epistemic labour between *inside* strategies, which mitigate the effects of reliabilist epistemic vices by implementing better reasoning processes, and *outside* strategies, which mitigate the effects of responsibilist epistemic vices by modifying the environment where biasing vices proliferate.

We argue that the complex architecture on which Bland's coordinated approach is grounded appears to lose some stability once we analyze more closely its pillars. We shall concentrate our attention on the notion of reliabilist and responsibilist epistemic virtues that the approach should foster as well as on the educational implications of Bland's view.

Consider first Bland's account of reliabilist epistemic virtues, according to which they are easier to cultivate (and more stable) than traditional responsibilist epistemic virtues. While this might be true of reliabilist epistemic virtues in general, it is not clear that these features apply to the reliabilist epistemic virtue that does much of the work on Bland's view, namely *sound mindware*. Mindware works like a cognitive faculty

and is in charge of our logical inferential capacities, statistical reasoning, and experimentation. Mindware counts as epistemically virtuous to the extent that it reliably produces accurate beliefs, but its acquisition and deployment are not as immediate and easy as our perceptual faculties and memory. If it is true that mindware can be trained internally through instruction and exercise, then it also involves a complex and varied set of competences, which presumably require time to be acquired and refined, much like the responsibilist epistemic virtues we can deploy to counter-biased cognition.

Furthermore, as Bland notes, for mindware to work effectively it is also necessary that the individual be aptly disposed and motivated to correct their posture toward their own reasoning processes. Besides marking a further difference between mindware and standard reliabilist epistemic virtues, this feature suggests that the acquisition of good mindware depends on the possession and correct deployment of responsibilist cognitive traits, which require instruction and habituation. Thus, it is far from clear that the key reliabilist epistemic virtue in Bland's view has an advantage over responsibilist epistemic virtues as regards how easy it is to cultivate the trait.

A further concern with the notion of sound mindware has to do with the responsibility that Bland associates to its correct deployment. Bland seems to think that through instruction and habituation one can learn how to reason in a way that mitigates one's proneness to cognitive biases and this, in turn, makes one responsible for failing to do so in the relevant situations. It strikes us as odd to concede that one can be held responsible for exercising (or failing to exercise) sound mindware. If responsibilist cognitive traits provide the necessary motivation for one to be aptly disposed towards discriminately exercising one's inferential capacities that form part of one's mindware, then the attribution of responsibility should target the enabling and motivating traits rather than the reliable ability (the sound mindware) itself.

Consider now the educational implications of Bland's approach. For Bland, the problem with standard epistemic virtue-based educational accounts is that they aim at fostering responsibilist epistemic virtues and thus appeal to internalist and individualistic strategies, which appear unable to counter biased cognition directly. The educational reform suggested by Bland is that epistemic virtue-based approaches include specific strategies of *cognitive outsourcing*. More specifically, these should be strategies that outsource the cognitive abilities through which students process available information. These strategies include helping students acknowledge the difficulties of personal debiasing (and the relevance of situational and environmental interventions in this regard) and highlighting the pros and cons of collective deliberation over individual deliberation as a way to counter-biased cognition.

Our concern with this proposal is that it is unclear that such cognitive outsourcing strategies involve any great departure from what an educational approach centred on responsibilist epistemic virtues would demand. For wouldn't the development of the responsibilist epistemic virtues in this educational context naturally go hand-in-hand with the cultivation of the kinds of strategies that Bland casts as "cognitive outsourcing"? Think, for example, of how the development of a responsibilist epistemic virtue like intellectual humility might dovetail with making individuals more aware of situations in which relying on their individual cognitive resources could be especially problematic. In short, it seems that what Bland is describing is less a critique of the educational role of standard responsibilist epistemic virtues than a credible description of what such a role should look like once fleshed out in a way that is suitably responsive to relevant empirical work on the amelioration of cognitive bias.¹

Note

- 1 As a concrete example of this point, consider the *Anteater Virtues* curriculum project at the University of California, Irvine, which is run by one of the present authors (DHP). This project is devoted to educating for the intellectual virtues, and thus for the responsibilist epistemic virtues, but it also includes, as part of this, practical guidance on how, for example, social media misinformation plays on one's cognitive biases, and how to guard against this. For a recent educational study of this project, see Orona and Pritchard (2021).

Reference

- Orona, G. A., & Pritchard, D. H. (2021). 'Inculcating curiosity: Pilot results of an online module to enhance undergraduate intellectual virtue', *Assessment & Evaluation in Higher Education*. DOI:10.1080/02602938.2021.1919988

1d Steven Bland's Response to Commentaries

Responses to Neil Levy's Commentary

Neil Levy recommends a strategy of reinterpreting seemingly irrational epistemic behaviour as rational responses to implicit testimony. On his view, many putative biases are the result of the *content* of the testimony we rely on, not the *practice* of relying on tacit testimony. Instead of engaging in debiasing, then, we would be better off improving the quality of the testimonial evidence available to agents under conditions of uncertainty.

This interpretation may work for some cases. The way we frame information often does convey our attitudes about it. I would add, however, that many of these attitudes are not epistemic, and therefore not all frames should be treated as testimonial evidence. I find Levy's treatment of the CRT more problematic. Levy classifies its questions as trick questions that implicate false answers, such as 10 cents in the ball and bat problem. He then asks: "Is it really less rational to be guided by the testimony they offer rather than to reject it and perform the arithmetic?" But the incorrect answer is also based on an arithmetical calculation, albeit at the level of intuition. When answering the bat and ball problem, many of us take \$1 away from \$1.10, and arrive at the answer of 10 cents. Thus, the questions on the CRT are trick questions not because they implicitly suggest incorrect answers, but because they reliably trigger incorrect operations at the level of intuition. It seems more plausible to think that our confidence in our answers is a function of the ease with which we perform the calculation, rather than a reliance on tacit testimonial evidence.

I should also take this opportunity to make a few clarifications. First, I do not count among the reliabilist inferential virtues only those forms of reasoning that conform to the content-blind norms of logic, probability theory, etc. There are heuristics, such as the recognition heuristic and take-the-best, whose use is ecologically rational under specific conditions (Gigerenzer 2008). Hence, to be inferentially virtuous, in the reliabilist sense, is to be *epistemically adaptive*, i.e., to routinely use the right mindware in the right circumstances. We can often cultivate epistemic

adaptability by means of targeted training, sustained practice, and quality feedback on our inferential performances.

I also think that responsibilist virtues (and vices) are context-sensitive: open-mindedness is epistemically deleterious in solitary circumstances, but not necessarily in dialogical conditions. But it is more difficult to cultivate and manifest these virtues *selectively*: we are unlikely to be open-minded in the absence of critical interlocutors. Effective training and copious practice and feedback are difficult to come by. So instead of attempting to adapt this type of behaviour to our cognitive environments, I have suggested that we are better off designing environments that better suit our typical behaviour. We might call this *epistemic accommodation*, rather than epistemic adaptability.

Response to Michel Croce & Duncan Pritchard's Commentary

Croce and Pritchard raise several valuable objections to my ameliorative framework for debiasing. I will answer as many as I can in the brief space I have.

First, they express doubts about the claim that the reliabilist virtues needed to mitigate cognitive biases are easier to cultivate than responsibilist virtues. I fully acknowledge that learning how to properly deploy sound mindware is a difficult undertaking that requires careful instruction, sustained practice, and quality feedback. My claim is only that responsibilist virtues, such as epistemic humility and intellectual vitality, are not as easily cultivated *by the same means*. On the flip side, they are *more easily* cultivated by strategies that involve situational interventions.

Second, they claim that "...the acquisition of good mindware depends on the possession and correct deployment of responsibilist cognitive traits, which require instruction and habituation." Thus, cultivating reliabilist virtues without the responsibilist virtues required to enable them seems like a fruitless pursuit. Here I should clarify my position: I believe that the consistent *deployment* of reliable mindware requires the manifestation of responsibilist virtues, but not the *acquisition* of mindware. Learning how to reason statistically is one thing; stifling intuitive responses in favour of statistical ones is another.¹ And while the latter task requires both the possession of reliable mindware and the manifestation of responsibilist virtues, our best ways of accomplishing these two things are distinct: inside strategies can impart mindware; outside strategies are more effective at promoting the responsibilist virtues needed to consistently deploy it.

Third, Croce & Pritchard see little difference between my proposal to teach strategies of cognitive outsourcing and the traditional program of cultivating responsibilist virtues in the classroom. They invite us to "Think, for example, of how the development of a responsibilist

epistemic virtue like intellectual humility might dovetail with making individuals more aware of situations in which relying on their individual cognitive resources could be especially problematic." I agree that intellectual humility is valuable in this respect, but I despair of the prospect of teaching *self-deployed techniques* of bolstering this trait. It seems clear that motivated reasoning and bias blindspot wreak havoc with our attempts to implement inside strategies, like considering the opposite, that target intellectual humility. Rather, what's needed is an *environment*, such as a forecasting tournament or a prediction market, that provides students with unambiguous feedback about the accuracy of their judgements. Designing, creating, and leveraging such environments is an outside-in strategy of cultivating intellectual humility by repeatedly exposing students to their own errors. Having developed some intellectual humility by this process, students may learn to seek out feedback mechanisms that help them to properly calibrate their confidence. In this way, cognitive outsourcing can promote traits that lead to more cognitive outsourcing. This is a view, I take it, that significantly departs from the traditional virtue theoretic approach of education by habituation.

Note

- 1 Stanovich (2011) distinguishes these achievements as exemplifying *crystallized* and *fluid* rationality, respectively.

References

- Gigerenzer, G. (2008). *Rationality for Mortals: How People Cope With Uncertainty*. Oxford: Oxford University Press.
- Stanovich, K.E. (2011). *Rationality and the Reflective Mind*. Oxford: Oxford University Press.

2 Attunement

On the Cognitive Virtues of Attention

Georgi Gardiner

1 Introduction

Attention matters. It influences our evidence, beliefs, knowledge, and understanding. It alters our conception of the world and our self-assessments, including whether we notice the limits of our understanding. Attunement is deeply tied to skills, values, and epistemic character. And, as I argue, it can be epistemically evaluated.

This chapter motivates three claims: Firstly, the normativity of attention is illuminated by virtue epistemology. Given deep connections between character and attention, it is fruitful to study the cognitive virtues of proper attunement (Section 2). Secondly, groups and collectives can possess virtues and vices of attunement (Section 3). Thirdly, attention is important for epistemology (see especially Section 4).¹ I highlight the social and ethical significance of attention for understanding disparate phenomena like media, social media, big tech, search engines, crime reporting, political polarisation, aims of political protest, sexual fantasising, and Lucifer's Fall. I use attentional normativity to undermine recent arguments for moral encroachment, the thesis that moral features of a belief can affect its epistemic justification. And I argue that putative cases of doxastic wrongdoing—that is, wronging someone by forming beliefs about them—might instead exemplify attentional wrongdoing or attentional vice. Highlighting the various interactions of epistemic and moral normativity can thus help defend purism, the view that whether a belief is epistemically justified depends solely on truth-relevant factors, such as evidence. Proper attunement is a deeply social phenomenon. We should be attuned to what matters; I suggest that the neologistic virtue of “wokeness” can be well-theorised as a virtue of proper attunement.

A recurrent theme is that beliefs, assertions, and various epistemic activities can be epistemically flawed even though all relevant propositions are true and well supported by evidence. This is because attentional patterns can distort, mislead, and misrepresent even when no claims are false. Relatedly, epistemic and communicative activities can be successful even though they neither uncover nor convey content. The conduct instead precipitates attentional shifts. Section 5 concludes by

emphasising the growing urgency of the epistemology of attention to understand the epistemic landscape of the internet age. Information is plentiful; we must assess information curation. Evaluative frameworks that are limited to whether propositions are true and evidentially supported are inadequate; a virtue epistemology of attention, I argue, provides valuable resources for this endeavour.

2 Proper Attunement Is a Virtue

This section posits cognitive virtues of attunement. I highlight central features of character virtues and vices and I show that attentional traits share these features. Note that my use of “attunement” differs from the psychologists’ sense of receptivity to and resonance with another person (Erksine, 1998). Nor do I simply mean having more or less focus, calibrated to the demands of one’s context and capacities, such as paying more attention when confronting difficult tasks and high stakes. And being attuned differs from having a “good attention span”. Proper attunement is paying attention to the right things in the right way, at the right time; being sensitive to significant features; and ignoring what should be ignored. It relates to questions of attention span, concentration, and quality of focus but, as will become clear, it is not exhausted by these.

Note too this chapter sidesteps contentious disputes about when attention is undue by using paradigm examples.² Fixating on Louvre bathroom signs rather than artworks, for example, typically exhibits improper attention. I do not develop principles for proper attention, but guidelines include that typically one should attend to central and illuminating features rather than peripheral details. Dwelling on risks is typically appropriate to the extent they can be managed, the outcome is severe or probable, or moral emotions are apt. It is typically inappropriate to dwell on farfetched or insignificant possibilities. Attentional patterns should reflect moral considerations and support aims like inquiry and happiness. Staring at Louvre bathrooms signs can be appropriate if you are redesigning them, for example, or are overwhelmed by crowds. Attentional normativity thus reflects manifold contextual features. This multiplicity and intricacy partially explains why virtue theory is well equipped to theorise proper attentional conduct.

2.1 *The Significance of Habits*

“I wonder whether my daughter gets enough iron”, thinks Ariana. “Vegan diets can be low in iron and Teagen is vegan”. Does Ariana pay too much attention to this question? We cannot tell. Our information is inadequate. To assess this we need to know, among other things, whether Ariana has reason to worry, what the evidence indicates, and whether iron consumption matters. We also need to know Ariana’s

broader thought patterns. It might be the first time Ariana has wondered this or she might think about it daily. Ariana illustrates that when assessing a person's attention, the locus of evaluation is often attentional patterns and habits, not instances.

Whether a chemical reaction is part of a living organism depends on its broader spatiotemporal context. This is not simply the epistemic claim that we cannot determine whether a process partially constitutes life without knowing what happened before, after, and around it. The claim is ontological: Whether the reaction is in fact part of life depends on those broader facts. Whether Ariana's thought constitutes improper attention is similarly dependent on diachronic features. Proper attunement depends on temporally extended cognitive conduct.

In some cases, a single instance of attention or inattention can be improper. Visually focusing on disfigurement, even fleetingly, can constitute improper attention, for example, regardless of broader attentional habits. And continuing a casual telephone conversation when a nearby stranger has just fallen from a pier constitutes inappropriate disregard. Even if you can't help them, their falling warrants attention. These examples illustrate that the loci of attentional normativity are not always attentional patterns. But typically attentional normativity depends on patterns and habits. Ariana's daughter is irked by a particular instance of her mum's wondering about her iron intake because her mum thinks of it too often. There is usually nothing wrong with isolated instances of wondering; an instance's badness stems from broader trajectories, habits, and traits. This accords with other virtue notions. Whether actions manifest virtues or vices typically depends on patterns of acting.³ In some cases, an action qualifies as virtuous or vicious regardless of the broader pattern, but typically an instance of, say, not donating money is not by itself a mistake; the mistake resides in miserly habits.

2.2 *Interrelated Facets of Agency*

Questions about proper attunement arise for diverse aspects of agency, including perceptual attention, occurrent beliefs, and what a person wonders, daydreams, cogitates, questions, doubts, and dismisses. Attention determines which possibilities a person takes seriously and which environmental features they are sensitive to, monitor for, and neglect. Attunement affects—and is partly constituted by—patterns of inquiry, communication, and forgetting. The heterogeneity of agential capacities that facilitate and govern attention indicates deep links between attentional normativity and cognitive character. To see this, contrast attunement with aspects of cognitive normativity that are more plausibly severed from character. Whether a belief constitutes knowledge, for example, is plausibly theorised by the characterological resources of virtue epistemology or by rival non-characterological frameworks like

evidentialism or coherentism. Proper attention, by contrast, seems inherently, ineliminably linked to character and thus the distinctive purview of virtue theory. In other words, virtue theory is well-positioned to limn attentional normativity because attention is essentially interlaced with heterogeneous but integrated parts of the character.

Some attentional features of a person are automatic, such as finding sudden noises salient. Some are habitual or associative, such as associating Ozzy Osbourne with bats. Others are deliberate and controlled, such as when one focuses on maths. We can use interactions amongst these levels to enhance the virtues of attunement. That is, the fact that attentional traits arise at different levels partly explains their plasticity. We can consciously bring something to mind repeatedly, so it later becomes habitual or automatic. We can deliberately learn more about Osbourne to weaken the association with bats. Someone might be unattuned to signs of boredom in listeners and so consciously work to notice them. The signs consequently become more salient to her and no longer require deliberate attention. She hones her virtues of attunement. Advertisers exploit the relative ease of changing perceptual salience to shape deeper attentional habits. I return to this in Section 4. The interplay amongst cognitive levels, and their power in cultivating and corroding virtue, is characteristic of virtue.

Stemming from this heterogeneity, attentional patterns can be assessed in many ways. The patterns and underlying dispositions can be rational, reasonable, apt, judicious, useful, creative, misleading, distorting, unwise, harmful, or destructive. They can reflect well or poorly on the person's character. They can contribute to, and partially constitute, wisdom. Their effects can also be assessed, including morally, epistemically, and prudentially. This evaluative richness is indicative of characterological assessment, rather than rival evaluative frameworks, like those centred on consequences, reliability, or evidential probability. And virtue theory can help unify the various grounds, roles, and evaluations of attention.

2.3 Developmental Features: Education, Emotion, and Understanding

Attentional patterns can be improved or worsened over time, both deliberately and otherwise. Like other characterological dispositions, they are shaped by community and culture, including in ways that are difficult to notice. Denizens of a sports-loving culture will typically think of sport relatively often, for example, compared to people from other cultures. Sport-inspired metaphors and explanations will be cognitively accessible for them and, since attention is contrastive, they may think of other topics less. Sport's cognitive centrality can go unnoticed and unquestioned because it matches the person's cognitive cultural background.

Education affects attunement. Aims of education include steering attention and developing concentration capacities, and people are more disposed to notice something after learning about it. And, conversely, attunement affects education. Developing skills and expertise requires attending and we typically discover more about salient phenomena. Virtuosity and expertise are sometimes partly constituted by the ability to perform well without devoting attention to the task. But reaching this stage typically requires investing considerable attention. Theorising attention illuminates educational injustices because, for example, one can fail to notice educational lacunas unless those topics are made salient. A person might never learn west African history, for example, yet never notice this.⁴

A central aim of education is enhancing understanding. Understanding involves the apprehension of coherence-making connections amongst facts; the person who understands sees how things hang together.⁵ A topic's salience across diverse cognitive contexts fosters the grasping of explanatory connections and thereby enhances understanding. Suppose Lissa cares about climate change. Since emotions direct attention, global warming is thereby salient to Lissa more frequently. When other topics, such as food, gifts, education, or generational wealth inequality arise, Lissa is more disposed to concurrently consider climate change. The topics cognitively coappear. These attentional patterns help Lissa forge explanatory links—whether accurate or erroneous—between climate change and these other topics. Since appreciating such connections is constitutive of understanding, Lissa's attentional patterns aid understanding.

Love, guilt, and trauma are powerful influences on attention. They thus can aid understanding and can distort. The conception of understanding sketched here helps explain why. The person seized by love, guilt, or trauma has their attention directed towards a topic across varied cognitive contexts, which causes them to forge novel connections, whether insightful or illusive, between that topic and others. Phobias and hatred distort a person's understanding, as the attentional forces of emotion help forge the links characteristic of understanding, but inaptly.⁶

Standpoint epistemology emphasises that occupying marginalised social positions affords distinctive epistemic benefits.⁷ The potency of attention for enhancing understanding helps illuminate how. The marginalised person's attention is drawn to the same topic, such as police brutality or wealth inequality, in diverse social and cognitive contexts. This process helps forge coherence-making connections amongst apparently disparate topics. Indeed the epistemology of attention suggests an epistemic value that arises from occupying particular social circumstances and cannot be acquired by testimony. Attentional patterns and dispositions are diachronic. If having appropriate attentional patterns

or traits has epistemic value, this value may ineliminably require the unfolding of time. It is not something that can be gained second-hand by, for example, deferring to the marginalised person.⁸

2.4 Feedback Loops: Values, Character, and Attention

Attention is integrated with other features of a person's character and values. Attentional patterns manifest, shape, and reveal epistemic and moral character.⁹ Suppose Carrie tends to notice the expensiveness of people's outfits. Recall attention is contrastive. Character can be revealed by attending to a person's clothing instead of, say, their wit or sadness. And, as with education, the connection is bidirectional. Character and attentional patterns form a feedback loop. Habitually noticing clothing leads to further sartorial beliefs, inferences, and predictions. Carrie's sensitivity to clothing can thus reveal and strengthen her social acuity. She will perceive patterns—perceptively or spuriously—between clothing and personality or social status. She regards clothing as significant *because* she notices it. Sartorial choices increasingly feature in Carrie's evidentiary and explanatory inferences. Ignoring a person's clothing could accordingly seem like, and indeed *become*, the epistemic error of neglecting evidence.

Attention is potent. Clothing choice does not merely *seem* more evidentially and socially important if people attend to it. Mere attention can render something important, which fuels further attention. Attentional feedback loops can be seen, for example, in attention to celebrities' political opinions. Those opinions matter when, and because, people attend to them.¹⁰ This illustrates how attentional patterns shape what people *should* pay attention to. Attention snowballs.

Sometimes attentional feedback loops are simply distorting. A person frequently exposed to news stories about violent crimes committed by immigrants will likely overestimate the prevalence of such crime. They may foster increasing resentment of immigration and erroneously centre such crime in their explanations of other social maladies. The constant attention restructures their values, character, evidence, and thought patterns.¹¹

Attentional patterns can either accord with or conflict with a person's broader character, values, and commitments. This too is characteristic of virtue-relevant conduct. Suppose Arthur disproportionately notices whether women are slender, for example. Theorising disproportionate attention lies beyond the scope of this chapter, but suppose the pattern far outweighs Arthur's attention to men's figures and the actual importance of physique. Arthur's attentional disposition can clash with his broader feminist commitments. But attentional patterns can qualify as "out of character" only to a point. Absent special explanation, a person cannot uncharacteristically be late on most occasions. It is instead

characteristic; they lack punctuality. Similarly, a person cannot uncharacteristically daydream about fame if those thoughts are constant and continual. Our habits become us.

2.5 Excellences of Character: Attunement and Other Virtues

Possessing good attentional traits—the virtues of attunement—is not simply a matter of following clear, determinate rules. Attunement requires responsiveness to subtle, hidden, abstruse, competing, or multi-faceted features whilst navigating disparate, complex, changing contexts. And so evaluating attunement requires virtue notions like excellence, competence, discernment, judiciousness, and intellectual dexterousness, which indicates attentional traits are characterological.

Attention plays cardinal roles in possessing and employing other virtues.¹² This may include, for example, modesty as not dwelling on one's good qualities and gratitude as focusing on one's good fortune. Virtuous forgiveness involves not dwelling on being wronged. Perhaps virtuous friendship includes focusing on friends' admirable qualities rather than their vices. (Note I do not endorse this claim because good friends attend to vices to help friends improve and disproportionate attention can be distorting, even when all the beliefs are true.)

Indeed proper attunement facilitates and guides other virtues. Attending is a prerequisite for properly assessing and responding to almost every context. Virtuous friendship requires understanding and helping friends, for example, which requires perceiving and appreciating their foibles, strengths, values, challenges, and so on. It requires noticing patterns, including ones they may themselves overlook. Suppose someone often complains about their job and starts pining for their hometown. A good friend might appreciate the significance of these apparently unrelated facts and be attuned to the connection: Their friend is considering—perhaps subconsciously—moving home. But this requires noticing subtleties. Similarly, being a virtuous teacher, researcher, or nurse requires attunement to features of the professional environment. Perhaps, then, attentional virtues are meta-virtues, prerequisites for, and constituents of, other virtues.

3 Collective Virtues of Attention

Section 2 argued that proper attunement is fruitfully cast as a cognitive virtue. This section posits that groups can possess attentional virtues and vices. Since this chapter already covers many topics, I sidestep contentious discussions about the nature of collective agency and virtue.¹³ I do this by focusing on less controversial examples. Readers who doubt collectives can have cognitive virtues are unlikely to be convinced by what follows, but I hope they find something of value in the chapter,

nonetheless, in its attention to the epistemology of attention. This side-stepping is itself an exercise in directing attention. I hope to avoid the mires of theorising group agency because I have different aims, namely foregrounding normative contours of attention and suggesting the reasons for conceiving of attunement as a virtue indicate that attentional virtues and vices are attributable to groups, institutions, and perhaps societies.

3.1 *Socially Distributed Attention*

Section 2 argued that whether attention is appropriate can depend on broader attentional patterns. In the case of Teagen's mother, the loci of normativity are patterns and traits, not any particular attentional instance. Attentional patterns also emerge across people. Suppose Teagen mentions her veganism on Facebook. If almost everyone who sees Teagen's post wonders whether Teagen receives enough iron, this constitutes disproportionate attention. As with her mother, plausibly this excess is not located in individual instances. It emerges from the aggregate.

For individual Facebook friends—or some, many, or most of them—the attentional instance is plausibly not inappropriate. There is typically nothing wrong with an individual's sometimes wondering about veganism and iron deficiency. Unlike with belief, there is considerable latitude in what we may wonder.¹⁴ And there are reasons to wonder. It is not outlandish that a vegan has low iron. Non-heme, plant-based iron is relatively hard to absorb. Yet the resulting pattern of socially distributed attention is disproportionate. Society unduly fixates on putative inadequacies of vegan nutrition, especially given that vegans are typically nutritionally healthier and better informed than non-vegans and given the relative neglect of health risks of non-veganism.¹⁵

Detractors might insist that emergent socially-distributed attentional patterns cannot be improper unless the individuals' attention is improper. They might argue individual Facebook friends are being nosy, for example. Perhaps Teagen's nutrition is not their business because they won't affect it or because their attention stems from ignorance. In response: Firstly, it is unduly judgemental to condemn these Facebook friends. Many exhibit concern for Teagen. We are free to wonder about all kinds of things, including topics we cannot control and lack expertise about. Undisciplined wonderings and considering diverse objects of passing thought are essential for creativity, curiosity, and enhancing understanding. And some Facebook friends might worry precisely because they understand nutrition. Secondly, readers can themselves devise an example they find compelling. The structure is that some, most, or all individuals do not exhibit a flaw in their attentional pattern by noticing or considering something, but the aggregate pattern is disproportionate.

3.2 *Group Attentional Traits*

Teagen's Facebook friends are not a group agent or promising candidate for attributions of group-level virtue or vice. The example simply illustrates how attentional patterns arise amongst people synchronically, in addition to intrapersonally diachronically. To investigate group-level virtues of attunement, it will be helpful to consider a paradigm group agent, such as a small deliberative decision-making group.

InvestyGate. A six-person investment group, InvestyGate, discusses whether to invest in AmaRanch, a small amaranth farm in Kentucky. It looks like a safe, lucrative investment that will outperform rival investment opportunities. One group member, Wayne, raises a worry. If it rains torrentially throughout June, Wayne notes, the crop would be ruined. He is correct that a heavy June rainfall would render AmaRanch unprofitable. The investors discuss the possibility briefly. Kentucky rainfall is typically low and there is no special reason to worry this year. They move onto other considerations, such as whether there will be sufficient labour to harvest the autumn crop and whether recent increases in farro sales helps or hinders amaranth sales.

In this case, the group exhibited the cognitive virtue of attunement. They were sensitive to relevant considerations, paid them appropriate attention, and properly situated them in deliberations. They did not dwell on Wayne's worry.

There is latitude in proper attention. Given the unlikelihood of crop-destroying rainfall, it would probably have been perfectly reasonable for InvestyGate to have never considered it, just as they did not discuss other distant but possible risks, such as the farmer's negligently allowing her insurance to lapse before a crop-destroying fire. But, given this latitude, it was also perfectly reasonable to discuss it and move on.

Wayne's raising the concern can alter the normative landscape of attention. Once Wayne raises the possibility of excessive rainfall, perhaps the group should briefly discuss it and merely waving it aside would be negligent. His mentioning the possibility may constitute evidence that it is not farfetched and is attention-worthy. But had Wayne not raised the topic, the group can permissibly never consider it. If so, this illustrates a way that paying attention to a topic affects epistemic normativity.¹⁶

3.3 *Non-Summativism*

The group might exhibit virtues of well-calibrated attention even if one member fails to. Suppose Wayne doesn't let it drop. He researches weather trends and—even though the data show crop-destroying rainfall

is rare—he reraises the possibility. In some such cases, Wayne thereby allots the possibility of an unreasonable amount of attention and exhibits the epistemic vice of improper attunement. But the group itself can be nonetheless virtuous. Indeed, they can be better attuned in virtue of Wayne’s individually disproportionate attention. Suppose rain is a non-negligible risk that they would have disregarded but, because of Wayne’s fixation, they instead allot appropriate attention.

The group could instead exhibit group-level *improper* attunement. They could, with Wayne, dwell on the possibility during multiple meetings. They might disregard other factors, such as consumer trends and alternative investments. Wayne’s rainfall fixation impairs group attunement.

Institutions other than deliberative groups also exhibit attentional patterns that emerge at the collective level. Suppose many scientists research male-patterned heart disease, but very few research female-patterned heart disease, for example. This is disproportionate attention. An individual researcher might dedicate years to a particular kind of male-patterned heart disease. Plausibly her attention is not improper; it is not attentionally inappropriate or vicious for a scientist to be engrossed in specialised research. Research requires specialisation. But the scientific community’s pattern is improper.

A group might exhibit a well-balanced attentional distribution precisely because each group member is differently fixated. In some cases the individuals’ attentional dispositions are irrational, yet the group functions well in virtue of this skewed attentional distribution. This suggests the virtues of attunement are non-summative: A group can lack the collective virtue even though each member’s attention is virtuous. Suppose every doctor hired into a cardiology department virtuously specialises in an interesting and important kind of heart disease, for example, but the overall group wholly neglects female-patterned heart disease. Conversely, a group can possess attunement even when no member does. Suppose each InvestyGate member is unduly gripped by a different investment opportunity and neglect alternatives, but the group’s discussions thereby focus adroitly. Indeed the undue attention of individuals yields deep insights and ensures each prospect is discussed.

Note that proper attentional distributions do not suffice for virtue; the group may lack appropriate attentional dispositions and motivations, for example.¹⁷ Note too that skewed attentional distributions can be vicious even if the aggregate amount is appropriate. Suppose one member of InvestyGate focuses wholly on gender justice, to the exclusion of other topics, and no other member ever considers it. The group lacks virtuous attentional distribution, even if the amount of attention is unimpeachable. A social virtue epistemology of attention can limn these normative contours further.

3.4 *Levels of Attentional Infrastructure*

Section 2 argued that attentional traits exhibit features characteristic of virtue and vice. These features included bidirectional links between attunement and education, skills, values, other virtues, and other character traits, for example, and that being properly attuned requires navigating complex, nuanced, and changing features of one's cognitive environment. Collective attentional traits likewise exhibit these features, which suggests there are collective attentional virtues and vices. I will not sketch group examples of each property described in Section 2. I instead focus on just one, namely that attentional traits arise from heterogeneous, interlocking agential components. This property illuminates the attention-shaping power of big tech and social nudging (Sections 4.4 and 4.5).

Individuals' attentional dispositions arise from myriad aspects of agency, including perception, intellection, and imagination. Attentional patterns arise for features that are automatic, habitual, subconscious, associative, reflective, deliberate, and so on, and can be grounded in extended environmental conditions. We use connections amongst these facets to alter attentional habits. Collective attentional dispositions are similarly heterogeneous. InvestyGate's attention arises from, and is constituted by, group discussions, correspondence, conversations with outsiders, individuals' thought patterns, and so on. Rainfall estimates could appear in minutes, memos, agendas, action items, whiteboards, shared electronic folders, silly jokes, or offhand comments. Funding and person-hours can be assigned to researching rainfall. A consultant could be hired. These media direct and constitute attention.

Funding, space, and time are key attentional resources for most collective agents. But, of course, different collective agents have varied kinds of resources. Attentional resources can include a newspaper's column inches, an art gallery's wall space, and a university's campus layout. Is the library the focal point, for example, or the football stadium? And which departments are relegated to campus peripheries? Accessibility of information and similar abstract features of social infrastructure determine—and can constitute—attention. Search engine rankings are a potent attentional force.

Substrata of attention can remain relatively segregated. InvestyGate might discuss rainfall at length, for example, but not keep written records or perceive connections between weather and other topics. Perhaps rainfall is neglected when discussing a similar farm. Alternatively, they might integrate the topic. Their newfound sensitivity to the significance of rainfall on farming means those same concerns become salient in novel contexts. The group's attention helps them forge new connections. Such features constitute the group's cognitive character. They affect the group's epistemic position, including its understanding, judgement,

knowledge, evidence, sensitivity, reliability, expertise, and confidence. As with individual agents, if the group can be confident it hasn't overlooked important considerations, it is owing to integration amongst the manifold parts of attention infrastructure. These attentional features underwrite the group's attentional virtues and vices.

3.5 *Group Action and Character*

As with individual attention, group attention is deeply linked to education, values, and character. These links are multi-directional and have feedback loops. A group's values shape its attentional patterns, for example, which in turn shape its values. And, as with individuals' attentional traits, an instance or pattern of attention can conflict with the group's broader values and character. InvestyGate might become uncharacteristically fixated on rainfall, for example, whereas typically they adeptly proportion attention.

Detractors might doubt the possibility of divergence between a group's values and its actions, including its attentional patterns. According to this objection, whilst an individual's values can diverge from her actions, a corporation's values are wholly determined by its actions. There is no space for the disparity to arise. If correct, this closes the gap between group attentional patterns and putative character traits. This threatens a virtue-theoretic treatment of group attentional dispositions because, if correct, groups cannot have attentional virtues and vices, but merely attentional patterns.¹⁸

In response, I concede that—compared to individuals—groups might be relatively constrained in their capacity to act out of character. Whereas an individual's valuing might be constituted by their history, emotions, motivations, hopes, thoughts, and other psychological and somatic states, an institution's valuing is more fully determined by its actions. But although slimmer, there is nonetheless a gap between a group's actions and its values and character. This gap is revealed by counterfactuals. Suppose researching female-patterned heart disease attracted accolades and career advancement. An ambitious research group, HeartLab, might devote considerable attention to the topic. But this behaviour does not determine HeartLab's values. If incentives were removed or better elsewhere, HeartLab's focus would shift. This illustrates a group's attentional patterns can diverge from its values and character.

Detractors might respond that HeartLab's attention and values do not diverge because its attentional patterns accord with a stable disposition to value careerism. In response, I grant HeartLab acts in accord with careerist values and traits, but they also—because of incentives—invest considerable attention towards women's health despite not caring about it.

Group's attentional patterns can also diverge from values and character simply because the group functions poorly. I thus hope to have motivated that collectives can possess attentional virtues and vices and these traits merit further investigation.

4 The Ethics and Politics of Attention

This section applies the virtue theoretic resources outlined above to highlight the importance of the epistemology of attention for understanding moral and social phenomena. A theme throughout the section is that attentional normativity requires epistemological frameworks beyond assessing whether claims are true and supported by evidence.

4.1 *True Yet Distorting*

Attention can render information misleading or inapt even when every claim is true. For audio, visual, print, and internet-based media, attentional patterns are determined and constituted by features like colour, shape, size, font, links, layout, volume, time, and motion. Suppose a news site reports crime. The reports are accurate and they carefully reflect overall ratios of crimes committed by immigrants and non-immigrants. But the site foregrounds the former; those reports are higher and have larger fonts. Since the website's claims are true and the ratios proportionate, criticising this news site requires evaluative frameworks from the epistemology of attention.

An organisation can hide detrimental information by not releasing it. But sometimes releasing information is legally required because of, for example, litigation or transparency laws. The organisation can instead bury the information within a camouflaging informational cacophony. This practice—known as “document dumping”—is similar to politicians strategically releasing damaging information during busy news cycles and on Friday afternoons. The released information is within the recipient's view, but not their grasp. It is difficult to criticise this epistemic misconduct using epistemological frameworks that focus on whether claims and assertions are true and evidentially justified. Assessing such practices requires an epistemology of attention.

Search engine results do not present themselves as accurate or inaccurate, but rather ordered by relevance or anticipated value to the searcher. Resulting rankings can distort even if every search result and linked website contains only accurate claims. Suppose, for example, that Googling “Guantánamo Bay” produces results about holiday accommodations above results about the infamous detention centre. The relative spatial location is epistemically inapt because it reflects reality poorly. Social virtue epistemology of attention offers resources for evaluating the power and influence of big tech companies.¹⁹

4.2 Political Polarisation Despite Full Agreement

The epistemology of attention illuminates political polarisation. Two individuals or groups could have similar beliefs and credences, and yet polarise as a matter of emphasis and attentional patterns. One worldview foregrounds crime and the other poverty, for example, in thought patterns, including associative dispositions, inferential habits, and time spent on topics. Attention-based polarisation is not captured by existing attempts to taxonomise and understand the epistemology of political polarisation, such as Talisse (2021), because attention-based, epistemic polarisation can happen even when people have identical belief content and confidence levels. This polarisation is insidious, entrenched, and hard to theorise and remedy because it is difficult to notice, measure, test, and criticise divergent attentional patterns, as compared to divergent beliefs.

4.3 Attentional Vice, Attentional Wronging, and Moral Encroachment

Attention has moral import. Suppose InvestyGate's Wayne continually raises the question of whether InvestyGate's secretary is embezzling funds, for example, despite lacking evidence. This can morally wrong her. But, importantly, this wrong hinges on Wayne's attentional patterns, not his beliefs. He might, after all, believe she is innocent. "Merely" asking questions can cause bad epistemic and practical downstream effects, such as when "mere" question-raising stoked early vaccine scepticism. But plausibly question-asking can *itself* constitute attentional wrongdoing or flawed attentional conduct, even aside from downstream causal effects.

Recently theorists have argued that beliefs can morally wrong a person.²⁰ Moral encroachment holds that moral features of a belief can affect its epistemic justification. Some adherents also endorse the principle of doxastic wrongdoing, which holds that beliefs about a person can morally wrong them, even when those beliefs are supported by good evidence, in virtue of the belief's content. These views are motivated by examples of, for example, believing someone is staff based on their race. They challenge the "purist" orthodoxy that, roughly speaking, whether a belief is epistemically justified depends only on evidence and other truth-relevant factors. Focusing on attentional normativity helps rebut arguments for moral encroachment and doxastic wrongdoing.

Rima Basu (2021) motivates moral encroachment and doxastic wrongdoing by noting that "We care how we feature in the thoughts of other people and we want to be regarded in their thoughts in the right way". But thinking isn't limited to belief. It includes characterological features, such as patterns of attention and inquiry. Emphasising this helps reconcile Basu's contention that thoughts can wrong with the purist claim that

a belief's epistemic justification doesn't depend on practical or moral factors. That is, we can countenance many epistemic and moral cognitive missteps without denying purism. Pointing to these missteps can defend purism because many putative examples of moral encroachment and doxastic wronging exemplify flawed *attentional* conduct and character, rather than flawed *beliefs*.

Moral encroachment is typically motivated via vignettes of, for instance, racial profiling. Vignette protagonists can commit multiple errors concurrently, and so identifying their purist-compatible errors is consistent with the vignettes also illustrating moral encroachment. But the point is dialectical: We can respond to arguments for moral encroachment by diagnosing flaws exhibited by the protagonists that are compatible with purism. This appeal to attentional normativity exemplifies how focusing on ethical and epistemic questions beyond whether a particular belief is justified by current evidence helps defend purism against challenges from moral encroachment. Purism is a narrow claim about the epistemic justification of individual beliefs at a time; it is thus consistent with myriad significant interactions between ethical and epistemic normativity. These rich normative ecotones can explain the vignettes that motivate moral encroachment within a purist framework.²¹ And virtue theory creates space to identify flaws and other places for improvement without decrying it a "wrong", that is, wholly prohibited conduct.

Attentional wronging, assuming it's possible, might manifest in various ways. Suppose two InvestyGate colleagues were formerly married and one frequently mentions this during meetings. Within InvestyGate, the fact is common knowledge, entailed by background evidence, and sometimes—for recusals, for example—relevant. Yet drawing gratuitous attention to this common knowledge might constitute an attentional wrong. Attentional normativity must distinguish, of course, amongst merely attending to a topic, deliberately drawing your own attention to it, and steering other people's attention. Some cognitive conduct, such as indulging in inappropriate sexual or violent fantasies, might qualify as attentional wrongs or flawed attentional conduct even if never disclosed to others. Perhaps sexually fantasising about a person who clearly doesn't want you to can manifest flawed character, for example. But one must be cautious about morally assessing thoughts. Sexual harassment is often glossed as "unwanted sexual attention", but the term "attention" is ambiguous between behavioural and mental conduct. This raises the spectre of sexual harassment merely by thinking about somebody.²² These potential sources of cognitive wronging are not doxastic wronging; the central phenomena are not belief.

Indeed plausibly attentional normativity offers a unified explanation of various (putative) wrongs or flaws of several doxastic and non-doxastic cognitive propositional attitudes and conduct, such as hoping, fearing, expecting, suspecting, doubting, imagining, daydreaming, ignoring,

forgetting, overlooking, and believing that p. Suppose a parent fervently hopes their child becomes a talented pianist, for example.²³ If the hope itself can wrong the child, or be flawed, it could be because the parent attends to the ambition too much. The connections between a person's values and attentional dispositions, discussed in Section 2, help explain why people care what others attend to.²⁴

4.4 Agenda Setting, Big Tech, and the Social Infrastructures of Attention

This section motivates two claims: If attentional patterns are invisible, the underlying attentional *infrastructure* is even more so. And big tech companies yield both attention-shaping powers. That is, we recognise that big tech determines what people pay attention to—the topics of attention—but big tech also sculpts the underlying attentional infrastructure that determines these attentional dispositions.

Wayne from InvestyGate illustrates that an individual can influence the topics of group attention. But Wayne can also shape his group's foundational attentional infrastructure. Suppose InvestyGate's meetings lacked agendas and Wayne introduced that structural resource for guiding attention, for example. Agenda setting is a powerful role. The minute-writer steers resources corresponding to group memory. Agenda setting determines group attention. Attention may be more foundational than values and judgements, since it determines what topics one has judgements about. Like many powerful roles, agenda setting can be invisible. Contours of attention—like the air we breathe—are hard to notice. Like other foundational aspects of cognition and social infrastructure, attentional patterns are typically noticed only when defective. Likewise, with breath.

Big tech shapes attentional patterns. Sometimes an attentional instance or pattern does not stem from a stable disposition, but instead reflects external forces. Suppose Teagen the vegan's Facebook comment was adjacent to adverts for iron supplements or an *Iron Man* movie on people's Facebook feeds. Her friends' wondering whether Teagen is iron deficient might stem from features of their cognitive environment—the salience of iron—rather than internal dispositions to wonder about vegan nutrition. But environments shape attentional dispositions and the proximity of iron supplement adverts to vegan content could be deliberate. If iron supplement adverts appear frequently, this fuels dispositions to think about micronutrition and associate iron deficiency with veganism. Advertisers exploit the relative ease of steering perceptual salience to shape deeper attentional character.²⁵

But more than this, big tech also shapes underlying attentional architecture. The Facebook corporation determined whether to have a separate “friends feed” and “current affairs feed” or to amalgamate them,

for example, and users barely notice or question this decision about the architecture of social attention. They are currently merged; one feed serves both functions. An epistemological assessment of this decision lies beyond the scope of this chapter, but the one-feed structure may fuel fake news, political polarisation, and the increasing dominance of social groups, rather than journalists, in shaping news exposure. Theorising the epistemology of attention helps distinguish these two distinct powers of big tech.

4.5 Progressive Nudges and the Aims of Political Protest

One can leverage different levels of attentional scaffolding to adjust attentional traits; we do this for individuals, groups, institutions, and societies. Female-patterned heart disease is under-researched. Individual scientists can begin to remedy this by asking questions about women's physiology at conferences, featuring female-patterned heart disease on course syllabi, or tweeting about the relative dearth of research. A medical association can direct attention by funding research, administering prizes, or hosting conferences. Journalists could foreground research on female-patterned heart disease and departments can encourage junior scientists by highlighting career benefits of this underexplored area. Shifting these various attentional levels alters overall attentional patterns.

Institutional features like newsletters, special issues, op-eds, and social media posts aim to shape knowledge and incentives. But they cannot be fully understood without focusing on their attentional aims. This is because many tweets and op-eds are not best understood as attempts to inform or incentivise: The audience either already knows, doesn't care, or won't remember the content. And more effective educational and motivational tools are available. The authors usually know all this. Yet tweeting (and similar) can nonetheless be effective because the author aims to influence attentional patterns, rather than inform. By creating instances of attention to female-patterned heart disease, individuals can help restructure overall attentional dispositions. The field thereby becomes more inclined towards noticing female-patterned heart disease and its research lacuna. The term "noticeboard" is telling; noticeboards often steer attention more effectively than they inform.

The virtue theoretic contours of attention illuminate the aims of protest. Chappell and Yetter-Chappell (2016) argue that inaction in the face of salient need is more monstrous than inaction concerning non-salient need.²⁶ They consider Peter Singer's influential comparison of a child drowning in a nearby pond and one starving abroad. Regardless of the overall choice worthiness of the two omissions, Chappell and Yetter-Chappell argue, inaction about the former exhibits worse moral character.

This insight illuminates the forces of protests. Protests can be effective. But how? They are typically not effective ways to educate or inform. Protest banners and chants might be humorous or build camaraderie, but they are feeble at transferring knowledge. And protests seem ineffective at directly affecting the observer's conative attitudes. Observers do not typically revise their motivations or beliefs about animal cruelty by seeing a protest.²⁷ Learning about animal cognition or talking with a friend is more effective at these aims. But protest is nonetheless effective: It directs attention. It reminds us that Guantanamo Detention Camp is still open, Washington DC lacks congressional representation, and polar bears face extinction. We already knew these things, but we weren't thinking about them. Roadblocks, celebrities, stunts, humorous signs, outlandish outfits, danger, and nudity can be effective protest techniques, not because they communicate relevant information but because they command attention. They attract media coverage, for example. Drawing attention to a need renders inaction more monstrous. People are motivated to not feel or appear monstrous. Thus the influence of attention on character virtues helps explain the efficacy of protest. "What-about-ism" in political discourse is similarly an exercise in directing attention. It directs attention away from an issue, which can make inaction seem—or be—less monstrous.

4.6 Wokeness, Liberation, and Attentional Injustice

Flaws in attentional distributions are often easily overlooked because default attentional patterns and infrastructure go unnoticed. Arthur may never notice that he disproportionately clocks whether women are slender, for example, partly because everyone around him does too. Attentional omissions—such as absences from an agenda or curricula and whose perspectives are missing from deliberations—are particularly hard to notice, diagnose, and remedy. And epistemic injustice can be caused by, and constituted by, attentional patterns of "tuning out" when some people, such as women, talk.²⁸

Unfair attentional distributions can arise when women are disproportionately expected to think of household demands or colleagues' emotional needs, for example, which allows men mental space to consider topics that advance their interests. The epistemology of attention illuminates epistemic contours of these disparities. Recall from Section 2 how attention, including background attentional tendencies, enhances understanding. Devoting background attention to interesting topics, rather than mundane household demands, provides cognitive advantages.²⁹ Proper attunement is vital to social justice, including as a liberatory virtue. It can be liberatory for women to pay less attention to their appearance, for example, and paying attention to marginalised groups serves and constitutes justice.

Wokeness is typically glossed as being aware of injustice.³⁰ Awareness has both informational and attentional components. It is not merely knowing; it requires “heeding” or bearing in mind. Wokeness is a virtue of attunement. By framing “wokeness” as largely an attentional trait, rather than centrally about one’s beliefs, one can countenance epistemic dimensions of wokeness without threatening evidentialist demands on belief. Gardiner (2018) emphasises that beliefs can serve justice whilst fully reflecting the evidence, but a conception of wokeness that centres on belief, rather than attention, risks requiring too much confidence in complex historical, economic, psychological, and social claims in the absence of requisite expertise and evidence. Attention-based conceptions of wokeness avoid this worry. And if wokeness centres on attention, rather than knowledge, it thereby avoids elitist values that cast ignorance and undereducation as moral flaws. Thirdly, interpreting wokeness as largely about attention, rather than belief or knowledge, accords well with early and influential recorded uses of “stay woke”, such as Lead Belly’s 1938 exhortation to Black travellers to Alabama to “watch out” and “be a little careful when they go along through there—best stay woke, keep their eyes open” and Erykah Badu’s contrasting “stay woke” with sleep, not ignorance, in her 2008 song “Master Teacher”.

This chapter argues that proper attunement is a social virtue because (i) it can be possessed by groups, collectives, institutions, and perhaps societies. And (ii) it is deeply affected by moral, interpersonal, and social factors. Social institutions, including especially big tech, should help cultivate attentional virtues. Social virtue epistemology can guide this endeavour.

5 The Devil Was Lost in the Details

This chapter investigates the cognitive virtues of attention for individuals and collectives. I argue that virtue theory provides a powerful framework for illuminating the complex, nuanced, diachronic, developmental, socially embedded contours of attentional normativity. Throughout the chapter, I highlighted the potency and importance of attention. Attention shifts the epistemic and moral landscape.

In closing, I highlight interconnections between two features of attentional normativity discussed in Section 4. Firstly, big tech companies play sizable roles in shaping attentional patterns and building the social infrastructures that underwrite those attentional patterns. They determine the Google rankings, for example, but also whether shopping, news, scholarship, and images have separate search results or not. Secondly, attentional patterns can be distorting even when all relevant claims are true and evidentially supported. Recall the website that disproportionately foregrounds crime committed by immigrants, for example. The articles can be wholly accurate—every claim is true—but

the disproportionate attention misleads and misrepresents. This is insidious because difficult to epistemologically criticise. Epistemological frameworks that focus only on whether claims are true and evidentially supported are inadequate. An epistemology of attention, by contrast, enables us to epistemologically assess the website because focusing on attention reveals a variety of epistemic errors that are consistent with the relevant claims being fully true and known.

These two features are importantly connected. In the internet age, vast swaths of information are available. Even when all the claims are true, information drowns in information. One can hide an object in plain sight by placing it in a messy room. In the information age, selecting, sorting, arranging, foregrounding, presenting, omitting, and contextualising information is paramount. These curatory epistemic virtues are indispensable.³¹ Epistemological frameworks that are limited to whether individual propositions are true and evidentially supported cannot epistemically assess big tech products, advise on navigating modern cognitive environs, or map normative contours of the social epistemic environment.

Lucifer's fall from grace raises a puzzle. Heaven was perfectly good, so how could Lucifer have erred? There was nothing imperfect for him to do, see, or desire. One response holds that Lucifer only focused on good things—there were no other—but his error was focusing on the less good things instead of the best things. Rather than contemplating the Divine, Lucifer was distracted by his own goodness.³² Analogous dangers lurk in our epistemic lives: Even *if* all our beliefs were true and well-founded, we could epistemically misstep by focusing on less attention-worthy things and being distracted by the paltry and peripheral.

The epistemic forces of attention can be insidious. It is difficult to notice, measure, evaluate, criticise, and remedy the patterns and infrastructure of attention, compared to, say, whether a claim is false or unsupported. And whilst many epistemological frameworks attempt the latter, there is a relative dearth of epistemological theorising aimed at the former. The informational cacophony of the internet age renders the epistemology of attention even more urgent. Attention demands attention.

Acknowledgements

This research greatly benefitted from a series of fruitful conversations with Rima Basu, Amy Flowerree, Liz Jackson, Renee Jorgensen, Jessie Munton, Steph Leary, Cat Saint-Croix, and Dennis Whitcomb. Thanks also to Spencer Atkins, Heather Battaly, Ning Fan, Hannah Fantuzzi, John Hardwig, Aeryn Longuevan, Ida Mullaart, Alasdair Murray, Wayne Riggs, Clerk Shaw, Josh Watson, and two classes at the University of Tennessee for helpful insights and suggestions. I am

particularly grateful to Mark Alfano, Jeroen de Ridder, Michael Ebling, Jon Garthoff, Paige Greene, Linh Mac, and Jacob Smith for comments on earlier drafts. Finally, many thanks to Adam Carter and Sandy Goldberg for their insightful responses, published in this same volume. This research was supported by a Graduate Research Award from the University of Tennessee Graduate School and by an ACLS Fellowship from the American Council of Learned Societies.

Notes

- 1 The epistemology of attention is strikingly underexplored within analytic epistemology. In September 2020 a Google Scholar search for “epistemology of attention” generated just 21 results. Most of these were poetry, education, or media studies, rather than philosophy. The rest were the philosophy and psychology of perception (Mole et al., 2011; Watzl 2011). Watzl (2017, 5) describes a similar dearth of research. Research in economics, media studies, communications, informatics, and psychology reveals the importance of attention (Lanham 2006). And Buddhist, Islamic, and Confucian traditions foreground attention (Ganeri 2017). Analytic philosophy appeals to the importance of attention in, for example, proper moral conduct and aesthetic appreciation (Herman 1993, esp. 73–93; Murdoch 2003, 16–36; Brewer 2009; Korsmeyer 2011, Todd 2014). The lacuna is epistemological theories of attention. Much of the analytic epistemology of attention is relatively new, such as Hookway (2003), Fairweather and Montemeyer (2017), Siegel (2017, n.d.), and Munton (2021). See also references in later footnotes.
- 2 For tractability I focus on appropriate attentional distributions, rather than attentional manner. Both are important. One should be sensitive to a person’s disability, for example, but not transfixed. These can involve similar attentional magnitudes, but a different manner. I also sidestep whether proper attunement is one unified virtue or a cluster. This depends on the individuation conditions of virtues.
- 3 Herman (2007, 1993) and Garthoff (2015).
- 4 I am grateful to Mark Alfano and Zoe Johnson King for helpful discussions on these topics.
- 5 See Kvanvig (2003), Elgin (2006), and Gardiner (2012).
- 6 On the epistemology of emotion directing attention, see Elgin (1999, 146–169), Goldie (2004), and Brady (2010, 2013). On affect directing aesthetic attention, see Korsmeyer (2011) and Todd (2014).
- 7 Note that standpoint epistemology is characterised by stronger claims and standpoint epistemologists offer various explanations for the epistemic benefits of social marginalisation (Toole 2019, 2022; Saint-Croix 2020). Thanks to Catherine Elgin, Amy Flowerree, Renee Jorgensen, and Cat Saint-Croix for conversations on these topics.
- 8 On some views, character traits are merely dispositional and do not require time unfolding.
- 9 Scanlon (1998, 39ff.) and Bommarito (2013).
- 10 See also Archer et al. (2020) on celebrity political opinions.
- 11 Munton (2021) and Watzl (2017).
- 12 For insightful discussion, see Chappell and Yetter-Chappell (2016) and Bommarito (2013).
- 13 Lahroodi (2018) discusses how requirements on collective agency and virtue affect collective virtue attributions.

- 14 There are epistemological normative conditions on non-doxastic attention. That is, wondering, daydreaming, imagining, considering, hoping, and similar can be epistemically inappropriate. But doxastic attitudes—belief, doubt, certainty, suspension—are more epistemically constrained.
- 15 Non-vegans face higher risks of cancer and cardio-vascular disease and, according to the National Institutes of Health, about 65% of the global population develops lactose intolerance (Orenstein 2017). Presumably the undue attention to putative risks of veganism is partly fuelled by animal farming industries and individuals' moral unease about their own omnivorism.
- 16 Elsewhere I suggest that mere attention can render error possibilities relevant and so undermine knowledge. Gardiner (2021-b) argues this is an epistemic mechanism of gaslighting, conspiracy theories, and other epistemic injustice. Gardiner (2021-a) questions whether society-wide attention to the possibilities that rape accusers are lying can render those error possibilities relevant. See also David Lewis's (1996) "rule of attention".
- 17 Thanks to Ning Fan for raising this issue.
- 18 See also Siegel (n.d.). I am grateful to Dennis Whitcomb for discussion on these topics.
- 19 I am grateful to Jessie Munton for conversations on these topics. See also Alfano and Skorburg (2018).
- 20 On doxastic wronging, see Basu and Schroeder (2019) and Basu (2021). On moral encroachment, see Bolinger (2020a, 2020b) and Gardiner (2018, 2021-b).
- 21 See Gardiner (2018, 2021-b) for further discussion of this dialectic.
- 22 Perhaps incessantly thinking about another person can attentionally wrong them. See Gardiner (forthcoming-b) on the obsessive attentional patterns that characterise limerence.
- 23 Basu (forthcoming) discusses these kinds of cases. To help isolate the normativity of the hope itself, one might assume the hope neither causes nor arises from poor parental behaviour.
- 24 This final sentence is ambiguous; I endorse both readings. This discussion benefitted from a series of conversations on intersections of ethics and epistemology with Rima Basu, Renee Jorgensen, Amy Flowerree, Liz Jackson, Steph Leary, and Cat Saint-Croix. I am grateful.
- 25 On the attention economy, see Lanham (2006), Wu (2017), and Williams (2018). Thanks to Mark Alfano and Dennis Whitcomb for discussion.
- 26 Mullaart (n.d.) notes that salience is observer-dependent. Theorists should avoid the consequence that, for example, an individual who is more attuned to others' distress because of her own traumatic history is thereby more monstrous for inaction than someone who simply fails to notice.
- 27 Protests might effectively shift *protestors'* attitudes, since people care more about subjects they have already invested in.
- 28 I am grateful to Adam Carter for this suggestion.
- 29 On attention and epistemic injustice, Gardiner (2021-b) examines the role of attention in gaslighting and conspiracy theories. I argue that focusing on remote error possibilities can render them relevant and so undermine rational belief. Similarly, Gardiner (2021-a) examines how focusing on the chance that a rape accuser is lying might render the error possibility relevant.
- 30 On the term's history see Pulliam-Moore (2016) and Romano (2020). For philosophical accounts of wokeness, see Basu (2019) and Atkins (2020). On attention, character traits and social justice, see Scheman (2017), Tanesini (2020, 59), Medina (2016), Gardiner (forthcoming-a), Whiteley (forthcoming), and Smith and Archer (2020).

- 31 The motto of University of Notre Dame's College of Arts and Letters is "Study everything. Do anything." This is bad advice.
- 32 This account of Lucifer's fall is an interpretation of Augustine (Burns 1988) and Anselm (Wood 2016, esp. 236–237). I am grateful to Josh Watson for perceiving this connection to Lucifer. This attention-based explanation of Lucifer's fall accords well with Bommarito (2013)'s attention-based treatment of pride: Lucifer was good. His downfall was paying too much attention to his goodness.

References

- Alfano, Mark and Joshua August Skorburg (2018) "Extended Knowledge, the Recognition Heuristic, and Epistemic Injustice" Duncan Pritchard, Jesper Kallestrup, Orestis Palermos and Adam Carter (eds.), *Extended Knowledge*. Oxford University Press, 239–256.
- Archer, Alfred, Amanda Cawston, Ben Matheson and Machteld Geuskens (2020) "Celebrity, Democracy, and Epistemic Power" *Perspectives on Politics* 18(1):27–42.
- Atkins, Spencer (2020) "Moral Encroachment, Wokeness, and the Epistemology of Holding" *Episteme*: 1–15. doi:10.1017/epi.2020.50
- Basu, Rima (2019) "Radical Moral Encroachment: The Moral Stakes of Racist Belief" *Philosophical Issues* 29:9–23.
- . (2021) "A Tale of Two Doctrines: Moral Encroachment and Doxastic Wronging" Jennifer Lackey (ed.), *Applied Epistemology*. Oxford University Press.
- . (forthcoming) "The Ethics of Expectations" *Oxford Studies in Normative Ethics*.
- Basu, Rima and Mark Schroeder (2019) "Doxastic Wronging" Kim and McGrath (eds.), *Pragmatic Encroachment in Epistemology*. Routledge, 181–205.
- Bolinger, Renee Jorgensen (2020a) "The Rational Impermissibility of Accepting (Some) Racial Generalizations" *Synthese* 197:2415–2431.
- . (2020b) "Varieties of Moral Encroachment" *Philosophical Perspectives* 34(1):5–26.
- Bommarito, Nicolas (2013) "Modesty as a Virtue of Attention" *Philosophical Review* 122(1):93–117.
- Brady, Michael (2010) "Virtue, Emotion, and Attention" *Metaphilosophy* 41(1–2):115–131.
- . (2013) *Emotional Insight: The Epistemic Role of Emotional Experience*. Oxford University Press.
- Brewer, Talbot (2009) *The Retrieval of Ethics*. Oxford University Press.
- Burns, J. Patout (1988) "Augustine on the Origin and Progress of Evil" *Journal of Religious Ethics* 16(1):9–27.
- Chappell, Richard Yetter and Helen Yetter-Chappell (2016) "Virtue and Salience" *Australasian Journal of Philosophy* 93(3):449–463.
- Elgin, Catherine (1999) *Considered Judgment*. Princeton University Press.
- . (2006) "From Knowledge to Understanding" S. Hetherington (ed.), *Epistemology Futures*. Oxford University Press, 199–215.
- Erksine, Richard G. (1998) "Attunement and Involvement: Therapeutic Responses to Relational Needs" *International Journal of Psychotherapy* 3:235–244.

- Fairweather, Abrol and Carlos Montemayor (2017) *Knowledge, Dexterity, and Attention*. Cambridge University Press.
- Ganeri, Jonardon (2017) *Attention, Not Self*. Oxford University Press.
- Gardiner, Georgi (2012) "Understanding, Integration, and Epistemic Value" *Acta Analytica* 27(2):163–181.
- . (2018) "Evidentialism and Moral Encroachment" Kevin McCain (ed.), *Believing in Accordance with the Evidence: New Essays on Evidentialism*. Springer, 169–95.
- . (2021-a) "Banal Skepticism and the Errors of Doubt: On Ephecticism about Rape Accusations" *Midwest Studies in Philosophy* 45:393–421.
- . (2021-b) "Relevance and Risk: Relevant Alternatives and the Epistemology of Risk" *Synthese* 199:481–511.
- . (forthcoming-a) "The Banality of Vice" Mark Alfano, Colin Klein and Jeroen de Ridder (eds.), *Social Virtue Epistemology*. Routledge.
- . (forthcoming-b) "We Forge the Conditions of Love" Carlos Montemayor & Abrol Fairweather (eds.), *Linguistic Luck: Essays in Anti-Luck Semantics*. Oxford University Press.
- Garthoff, Jon (2015) "The Salience of Moral Character" *Southern Journal of Philosophy* 53:178–195.
- Goldie, Peter (2004) "Emotion, Reason and Virtue" Dylan Evans and Pierre Cruse (eds.), *Emotion, Evolution, and Rationality*. Oxford University Press, 249–69.
- Herman, Barbara (1993) *The Practice of Moral Judgment*. Harvard University Press.
- . (2007) *Moral Literacy*. Harvard University Press.
- Hookway, Christopher (2003) "Affective States and Epistemic Immediacy" *Metaphilosophy* 34:78–96.
- Korsmeyer, Carolyn (2011) *Savoring Disgust*. Oxford University Press.
- Kvanvig, Jonathan (2003) *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press.
- Lahroodi, Reza (2018) "Virtue Epistemology and Collective Epistemology" Heather Battaly (ed.), *Routledge Handbook of Virtue Epistemology*. Routledge, 407–419.
- Lanham, Richard (2006) *Economics of Attention*. Chicago University Press.
- Lewis, David (1996) "Elusive Knowledge" *Australasian Journal of Philosophy* 74:549–567.
- Medina, José (2016) "Ignorance and Racial Insensitivity" R. Peels and M. Blaauw (eds.), *The Epistemic Dimensions of Ignorance*. Cambridge University Press, 178–201.
- Mole, Christopher, Declan Smithies and Wayne Wu (eds.) (2011) *Attention: Philosophical and Psychological Essays*. Oxford University Press.
- Mullaart, Ida (n.d.) "Salience, Hypervigilance, and Epistemic Injustice".
- Munton, Jessie (2021) "Prejudice as the Misattribution of Salience" *Analytic Philosophy*. doi:10.1111/phib.12250
- Murdoch, Iris (2003) *The Sovereignty of Good*. Routledge.
- Orenstein, Beth (2017) "Can You Become Lactose Intolerant Later in Life?" EverydayHealth.com.
- Pulliam-Moore, Charles (2016) "How 'Woke' Went from Black Activist Watchword to Teen Internet Slang" *Splinter News*. <https://splinternews.com/how-woke-went-from-black-activist-watchword-to-teen-int-1793853989>

- Romano, Aja (2020) "A History of Wokeness: Stay Woke. How a Black Activist Watchword got Co-Opted in the Culture War." *Vox*. <https://www.vox.com/culture/21437879/stay-woke-wokeness-history-origin-evolution-controversy>
- Saint-Croix, Catharine (2020) "Privilege and Superiority: Formal Tools for Standpoint Epistemology" *Res Philosophica* 97(4):489–524.
- Scanlon, Thomas (1998) *What We Owe to Each Other*. Harvard University Press.
- Scheman, Naomi (2017) "On Mattering" Giancarlo Marchetti and Sarin Marchetti (eds.), *Facts and Values: The Ethics and Metaphysics of Normativity*. Routledge.
- Siegel, Susanna (2017) *The Rationality of Perception*. Oxford University Press.
- . (n.d.) "Are There Norms of Salience?"
- Smith, Leonie and Alfred Archer (2020) "Epistemic Injustice and the Attention Economy" *Ethical Theory and Moral Practice* 23:777–795.
- Talisso, Robert (2021) "Problems of Polarization" Elizabeth Edenberg and Michael Hannon (eds.), *Political Epistemology*. Oxford University Press.
- Tanesini, Alessandra (2020) "Ignorance, Arrogance, and Privilege" Ian James Kidd, Heather Battaly, and Quassim Cassam (eds.), *Vice Epistemology and the Epistemology of Ignorance*. Routledge, 53–68.
- Todd, C. (2014) "Attention, Negative Valence, and Tragic Emotions" J. Levinson (ed.), *Suffering Art Gladly*. Palgrave Macmillan, 224–246.
- Toole, Briana (2019) "From Standpoint Epistemology to Epistemic Oppression" *Hypatia* 34(4):598–618.
- . (2022) "Demarginalizing Standpoint Epistemology" *Episteme* 19(1):47–65.
- Watzl, Sebastian (2011) "The Philosophical Significance of Attention" *Philosophy Compass* 6(10):722–733.
- . (2017) *Structuring Mind. The Nature of Attention and How It Shapes Consciousness*. Oxford University Press.
- Whiteley, Ella (forthcoming) "Harmful Salience Perspectives" S. Archer (ed.), *Salience*. Routledge.
- Williams, James (2018) *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge University Press.
- Wood, William (2016) "Anselm of Canterbury on the Fall of the Devil: The Hard Problem, the Harder Problem, and a New Formal Model of the First Sin" *Religious Studies* 52:223–245.
- Wu, Tim (2017) *The Attention Merchants*. Penguin Random House.

2b Commentary from J. Adam Carter

Reply to Gardiner on Virtues of Attention

Georgi Gardiner's "Virtues of Attention" sets out to do three main things: to (i) motivate the importance of attention for epistemological theorising; to (ii) argue that the normativity of attention is illuminated by virtue epistemology; and to (iii) highlight how the virtues of proper attention are plausibly conceived of as collective and institutional virtues, and not merely as individual virtues.

On point (i) I am in agreement. As far as I am aware, the most developed work on the epistemic significance of attention is found mostly in the philosophy of emotion,¹ and in the philosophy of perception,² rather than in mainstream epistemology; so Gardiner's contribution here is a welcome one. On point (iii) I am also in agreement. As Gardiner points out, "groups, institutions, and sets of people also exhibit attentional patterns" (Gardiner, forthcoming, X). Given that groups, institutions, and the like can plausibly exhibit attentional patterns,³ we should expect that the dispositions that give rise to them are (epistemically) better or worse.

While I am sympathetic to point (ii)—the claim that the normativity of attention is illuminated by virtue epistemology—I am less convinced that the tack Gardiner has taken in the chapter to establish this has done so convincingly. And so, from here on out—while I think Gardiner's chapter is rich and that it succeeds admirably in most of its aims—I am going to focus narrowly on (a) why I don't think Gardiner has really established *that* virtue epistemology illuminates the norms of attention; but—and this point is in a friendly spirit—I think that there is a very good case to make that virtue epistemology *can* illuminate (some) norms of attention, and I will explain, beyond what Gardiner has suggested, how I think it could potentially do so.

So what would it be to show that something (be it virtue epistemology, or anything else) "illuminates the norms of attention"? On the assumption that there are norms of attention (one I think Gardiner is right to make), such norms might be evaluative or prescriptive.⁴ Evaluative norms will say when some kind of attention pattern is good *qua* the kind

of thing it is.⁵ For example, “*Ceteris paribus*, attention that tracks valuable properties is better than otherwise”. Prescriptive norms of attention prescribe (permit or prohibit) attention patterns. For example, “Don’t focus on irrelevant details”. To illuminate either kind of norm would presumably involve identifying the source of the relevant normativity, or suggesting how we might go about identifying it. To this end, some questions we’d hope to answer are: *why* do evaluative norms of attention tell us that attention is better as such if it has certain properties rather than others? Relatedly: *why* do prescriptive norms of attention prescribe (or prohibit) some attention patterns but not others? What *explains* all of this?

For virtue epistemology to illuminate the normativity of attention in a substantial way, it would at minimum need to answer (or put us in a position to answer) these kinds of questions; put another way, it seems that appealing to virtue epistemology will *not* have illuminated the normativity of attention very well if it has left it mysterious, or just a brute fact, that the evaluative or prescriptive norms of attention are those that we take them to be.

Gardiner’s strategy for using virtue epistemology to illuminate the normativity of attention takes as a starting point “that proper attention seems inherently linked to character”. This seems true enough. Her strategy from here is to show that the “normativity of attention is illuminated by conceiving of being properly attuned as having cognitive virtue”, where proper attunement “is paying attention to the right things in the right way, at the right time; ignoring what should be ignored, and being sensitive to significant features”. But *what is it* that determines whether you’ve paid attention in the right way or the wrong way? Gardiner says her chapter is meant to be “relatively ecumenical about what determines whether attentional patterns are improper. Instead I focus on paradigm examples”.

Gardiner might be entirely right that being properly attuned involves having cognitive virtue, *and* that proper attunement requires paying attention in the “right way, at the right time”, etc. At the same time though, if the matter of what determines whether attentional patterns are improper is itself not explained (or such that we’ve been put in a better position to explain this), then there remains a straightforward sense in which the *normativity of attention* hasn’t really been illuminated in a substantial way yet, by virtue epistemology or otherwise.

The good news, though, is that I think virtue epistemology *can* help us to illuminate this; the tools of telic virtue epistemology⁶ offer just the kind of resources we’d need in order to better understand why (in short) proper attention is proper and improper attention is not. One convenient way to do this would be to construe the way we apportion our attention patterns as kinds of *attempts* in their own right. For example, suppose we intentionally aim to focus on a cognitive task T, whether it

be a simple task or a more complex task.⁷ With reference to this aim, we can then assess our apportioning our attention as *successful* or not, on the basis of whether the relevant aim is attained (or not). The success here might be accidental. Or the success might be due to the exercise of a disposition to proportion attention reliably (enough) when one aims to focus on T (or in relevantly similar cases). If issuing from such a disposition, the apportioning of attention would then be *competent*, regardless of whether it is successful. Finally, *apt* apportioning of attention will be not only successful and competent but successful *because* competent.

With reference to the above kind of picture, telic virtue epistemology offers the kind of framework within which we could potentially illuminate the evaluative normativity of attention, by giving us straightforward answers to how attention can be proper (or improper) along the three specific evaluative dimensions of success, competence, and aptness.

Is this fully satisfying? Not yet. After all one might aim to attend to some cognitive task, T, and then aptly apportion one's attention to T, when one ought *not* to have done so. For example, one might attend aptly to a trivial task. One's apportioning of attention in such a case is *still* apt, just as the executioner's movements may aptly attain their aim even when what is done is reprehensible, and so even if they should have had a different aim.

In telic virtue epistemology, it is acknowledged that there is a separate kind normativity that pertains to *which kinds of inquiries one should take up* in the first place. As Sosa puts it, this separate domain of normativity is the domain of "intellectual ethics".⁸ As I see it, the question of which tasks to turn your attention *to* falls in the area of intellectual ethics. Whether virtue epistemology (of any stripe) can illuminate those norms of attention that fall within intellectual ethics—and so those norms of attention stand outside of the kind of telic assessment applicable to aimed attempts as such—remains to be seen.

Notes

- 1 For example, according to Michael Brady (2010; 2013), the epistemic significance of emotions lies in the fact that they have the power to direct attention in the way that they do.
- 2 See, for example, Mole (2008; 2015) and Smithies (2011).
- 3 This is plausible both in a summative sense, though as well as in an inflationist or non-summative sense. For example, a jury might manifest attentional patterns by disproportionately deliberating about certain aspects of a case rather than others.
- 4 For discussion of this distinction, see, e.g., McHugh (2012, 22) and Simion, Kelp, and Ghijsen (2016, 384–386).
- 5 This evaluative "good" here is in Geach's (1956) sense that a sharp knife is a "good" knife, where "good" is a predicate modifier as opposed to a predicate.

- 6 The primary exponent of this view is Ernest Sosa. See, especially, his Sosa (2021). See also Carter (2021) for a recent variation on the view.
- 7 I am using a case featuring an intentional aim to simplify applying the model. The constitutive aim of a given attempt can also be set functionally. For example, as Sosa (2021, 25, fn. 12) notes, we can assess our implicit or “functional” beliefs for success, competence and aptness—those that guide behaviour below the surface of conscious reflection—not because a thinker intentionally aims at anything, but just because teleologically our perceptual systems aim at correctly representing our surroundings. For further discussion of functional and teleological assessment, see Sosa (2017, 71–72, 129–130, 152; 2021, 24–31, 52–58, 64, 110, 118).
- 8 See Sosa (2021, Ch. 2).

References

- Brady, Michael S. 2010. ‘Virtue, Emotion, and Attention’. *Metaphilosophy* 41 (1–2): 115–131.
- . 2013. *Emotional Insight: The Epistemic Role of Emotional Experience*. Oxford University Press.
- Carter, J. Adam. 2021. ‘De Minimis Normativism: A New Theory of Full Aptness’. *The Philosophical Quarterly* 71 (1): 16–36.
- Gardiner, Georgi. Forthcoming. ‘Attunement: On the Cognitive Virtues of Attention’. In *Social Virtue Epistemology*, edited by Mark Alfano, Jeroen de Ridder, and Colin Klein. Routledge.
- Geach, Peter T. 1956. ‘Good and Evil’. *Analysis* 17 (2): 33–42.
- McHugh, Conor. 2012. ‘The Truth Norm of Belief’. *Pacific Philosophical Quarterly* 93 (1): 8–30.
- Mole, Christopher. 2008. ‘Attention and Consciousness’. *Journal of Consciousness Studies* 15 (4): 86–104.
- . 2015. ‘Attention and Cognitive Penetration’. In *The Cognitive Penetrability of Perception*. Oxford University Press.
- Simion, Mona, Christoph Kelp, and Harmen Ghijsen. 2016. ‘Norms of Belief’. *Philosophical Issues* 26 (1): 374–392.
- Smithies, Declan. 2011. ‘Attention Is Rational-Access Consciousness’. *Attention: Philosophical and Psychological Essays*, 247–273.
- Sosa, Ernest. 2017. *Epistemology*. Princeton University Press.
- . 2021. *Epistemic Explanations: A Theory of Telic Normativity, and What It Explains*. Oxford University Press.

2c Commentary from S. Goldberg

“The Normativity of Attention: Characterological vs. Social”: Comments on Georgi Gardiner’s “Attunement: On the Cognitive Virtues of Attention”

Georgi Gardiner’s chapter advances the idea that social epistemology has much to gain by paying attention to attention. Her chapter aims to advance three main claims. The first is that “we should be attuned to the normative contours of attention”. The second is that when we do theorise about attention and its normative contours, we ought to conceive of these in “characterological, virtue-based” terms. And the third is that

the reasons for conceiving of attunement as a characterological, virtue-based notion suggest that attentional virtues and vices are also attributable to groups, institutions, and maybe even societies and other social phenomena.

(8)

There is much to admire about this chapter. For one thing, Gardiner’s first claim strikes me as both important and plausible, and her argument on this score will add much to the case that a select few others have made in their attempt to bring attention to the attention of epistemologists.¹ In addition, Gardiner’s argument on this score reinforces the case for virtue-theoretic approaches to epistemology: there can be little doubt that she is correct in thinking that a virtue theory not only accommodates but might be used to explain (at least some of) the normative dimensions of attention. And I should add, too, that her case for thinking of collectives as evaluable in light of the virtues and vices of attention is interesting and worth considering at greater length.

In this brief response, however, I will focus on the second of the three claims she makes: that when epistemologists theorise about the normative contours of attention, we ought to conceive of these in “characterological, virtue-based” terms. I want to suggest that there may well be cases in which the normative expectations on attention flow, not from

our conception of what makes for a flourishing (cognitive) life *per se*, but rather from the sort of social expectations that we have of one another when we are immersed (and play salient roles) in an epistemic community. While these expectations are assimilable within a virtue epistemology, they point to an additional source for the normative demands on attention: that source is not grounded in the value of a flourishing cognitive life.

I begin by acknowledging that at least some of the normative demands on attention *do* seem readily explicable in virtue-theoretic terms. Consider for example an injunction from the epistemology of testimony, to the effect that a good recipient of testimony ought to be attentive to signs of insincerity or incompetence. Failure to attend to such things when they are present increases the chances that one is taken in by false or otherwise unreliable testimony. Since being taken in in this way is not part of a flourishing cognitive life, we might take the demand to be attentive to such signs to be explicable in terms of its role in conducing to a flourishing cognitive life.²

However, not all of the normative demands on attention are readily explained in such terms. Some demands on attention flow from one's role in a community: lawyers ought to attend to (and remain on the lookout for) features of situations that bear on their clients' legal well-being, doctors ought to attend to (and remain on the lookout for) features of situations that bear on their patients' health, etc. To be sure, we might think that in each case there is such a thing as a flourishing cognitive life *qua* lawyer (or *qua* doctor, etc.). But I venture to suggest that cases are possible in which the demands themselves are neutral with respect to flourishing. These will be cases in which conforming to the normative demands on attention conduces neither to cognitive flourishing nor to cognitive languishing. Consider a person whose job it is to survey all of the parking meters in a given city to ensure that they are functioning properly, or a person with the responsibility of overseeing the production of high-quality ball-bearings at a local production plant. I would even speculate that there are possible cases in which conforming to the normative demands on attention might actually lead to cognitive languishing of a sort. Consider a therapist whose expertise concerns the relationship between cognitive decline and depression: she is tasked with being attentive to the signs of cognitive decline in her clients, but knowing what she does about the link with depression, the more attentive she is the more depressed she herself gets and the less motivated she is to continue her work (it is her stubborn sense of professional duty that keeps her going).

None of these possibilities suggest that Gardiner is incorrect about the significant contributions that virtue epistemology can make to our understanding of the normative demands on attention. Rather, they suggest that a virtue epistemology might not give us the complete story

about the range of those normative demands. And if I am right about this, then we can also conclude that the social dimensions of (the normative demands on) attention go beyond cases involving collectives and groups. In particular, we might think that our social life is itself a rich source of the normative demands on attention—a point whose proper explanation appears to require more than what is provided by virtue epistemology (at least as traditionally conceived).

Notes

- 1 See e.g. Schellenberg (2018), Siegel (2006, 2007), Watzl (2017), and Silins and Siegel (2019).
- 2 Arguably, this sort of idea is present in the virtue-theoretic approach to testimony endorsed by Fricker (2007).

References

- Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
- Schellenberg, S. (2018). *The Unity of Perception: Content, Consciousness, Evidence*. Oxford: Oxford University Press.
- Siegel, S. (2006). How does visual phenomenology constrain object-seeing? *Australasian Journal of Philosophy*, 84(3), 429–441.
- Siegel, S. (2007). How can we discover the contents of experience? *The Southern Journal of Philosophy*, 45(S1), 127–142.
- Silins, N., & Siegel, S. (2020). Attention and Perceptual Justification. In Adam Pautz & Daniel Stoljar (eds.), *Festschrift for Ned Block*. Cambridge, MA: MIT Press.
- Watzl, S. (2017). The philosophical significance of attention. *Philosophy Compass*, 6(10), 722–733.

2d Georgi Gardiner's Response to Commentaries

The Limits of Virtue?: Replies to Carter and Goldberg

Adam Carter and Sandy Goldberg both challenge my claim that the normativity of attention is well-illuminated by virtue epistemology. Carter suggests virtue epistemology may not address which attentional patterns and habits we should have. Goldberg points to demands on attention stemming from social roles, such as professions. Both criticisms are, I think, rooted in relatively narrow conceptions of virtue theory.

Virtue Reliabilism and Virtue Responsibilism

Carter contends that “To *illuminate* [the normativity of attention] would presumably involve identifying the *source* of the relevant normativity, or suggesting how we might go about identifying it” (emphasis mine). And, Carter continues, my proposal hasn't met this criterion. In response, firstly, this methodological requirement on illumination or explanation is too demanding.¹ One can illuminate or explain a phenomenon without providing a reductive explanation or tracing the phenomenon back to its ultimate source. One can illuminate via partial explanation or by occupying explanatory levels that aren't reductions to fundamental grounds. Just as claims from applied and normative ethics can be combined with metaethical and metaphysical claims about what fundamentally explains those claims, a virtue theory of attention is compatible with various accounts of why, fundamentally speaking, attentional patterns matter at all. Virtue theory might explain the source of attentional normativity, but this isn't required for virtue theory to illuminate attentional norms.

Carter claims that resources from virtue reliabilism explain the source of attentional normativity. On Ernest Sosa's view, knowledge is apt belief.² Beliefs are *Apt* when their *Accuracy* manifests *Adroitness*. Carter modifies this virtue reliabilist AAA framework to apply to attentional normativity. He suggests that one aims at attentional distributions towards tasks, and the resulting attentional distribution is proper when apt; that is, when attainment of the attempted attentional distribution manifests adroitness. This substitutes Sosa's *Accuracy* with *Attainment*,

because attentional distributions are not truth apt. Sosa's orthodox AAA framework applies to belief; Carter's adaptation of virtue reliabilism thereby exemplifies how theorising attentional normativity expands the concerns of epistemology beyond truth and belief.

But, Carter notes, this framework leaves unexplained which attentional patterns one should aim for. He concludes,

As I see it, the question of which tasks to turn your attention to falls in the area of intellectual ethics. Whether virtue epistemology (of any stripe) can illuminate those norms of attention that fall within intellectual ethics [...] remains to be seen.

I aver that virtue responsibilism, rather than virtue reliabilism, illuminates intellectual ethics. Virtue responsibilism is versatile, theorises multifaceted explananda, and features multiple dimensions of assessment. It considers social, moral, and contextual features, including motivations and personal development. Resources from virtue responsibilism and reliabilism might be fruitfully combined to yield a comprehensive framework for evaluating attentional traits and patterns. I lack space to explore this idea; I instead sketch three concerns about Carter's adaptation of virtue reliabilism's SSS framework.

Firstly, Carter's proposed framework is best suited to when a person (intentionally) aims at distinct, dissociable attentional distributions, such as during specific tasks. But these might be relatively marginal or abnormal cases. They are, anyway, a fraction of the target phenomenon. We need a framework for assessing automatic, default attentional habits and abilities as one navigates life. This includes general omissions, like not staring at physiological abnormalities, and sensitivity to complex situations' important features, such as a friend's capacity to notice sadness or a harried nurse's attunement to subtle symptoms. Similarly, we seek a framework for assessing overall life patterns and trajectories such as, for example, Greta Thunberg's attentional dedication to the climate catastrophe. But the AAA framework does not straightforwardly apply to these examples, not least because patterns and habits evolve over time, whether attention matches a given pattern is not binary, and attention is contrastive. The good friend doesn't aim to notice sadness, he is simply well attuned to emotions and conduct—or to other features, like the road he is driving on—as appropriate. This brings me to the second worry.

Attentional patterns are not sufficiently similar to true beliefs for Sosa's AAA framework to smoothly apply. Whether a belief is true is often binary and straightforward; the epistemic value of true belief is not wholly dependent on broader features of the person and context, and there is a relatively clear sense in which beliefs aim at truth. These features undergird Sosa's AAA treatment of the normativity of

true belief. Attentional traits and habits do not share these features. Whether an attentional pattern, habit, or trait matches an ideal can be complex and nuanced. It may match in some ways but not others, for example. And whether attentional patterns are valuable can depend wholly on social, moral, and contextual features, including the person's attitudes and motivations. And it is doubtful that we typically aim at attentional distributions, at least in an ordinary sense. These differences problematise adapting the AAA framework for attentional normativity.

Finally, Carter says “[the AAA framework offers] just the kind of resources we’d need in order to better understand why (in short) *proper attention is proper* and improper attention is not”. But it is unclear whether, absent an independent account of which attentional patterns are good, the virtue reliabilist AAA framework makes much headway on questions of *propriety*. For this, we need intellectual ethics.

Depths of Sociality

Goldberg emphasises demands on attention that stem from community roles. He notes these demands are “assimilable” within a virtue framework, so it is unclear how much we disagree. The crux of the disagreement—such as it is—is that Goldberg views these as “additional sources” of attentional demands, outside of virtue theory, because they are “not grounded in the value of a flourishing cognitive life”.

Goldberg appears to employ a relatively narrow conception of virtue theory, according to which the relevant attentional value or demand must directly contribute to the cognitive flourishing of that same individual. (See, *e.g.*, his testimonial illustration.) We might broaden this conception in several ways. Perhaps any flourishing qualifies, for example, not merely cognitive flourishing. This helps unify the ethics and epistemology of attention. Insofar as this is a correction, it is one I welcome; the attentional normativity interlaces ethical and epistemic considerations, and so is the domain of virtue theory, rather than virtue epistemology specifically. Secondly, the contribution need not be direct. Proper attunement in one’s employment can contribute to flourishing via salary, or pride in one’s work, for example, or via the mental health benefits of entering the “flow state”.

Thirdly, the relevant flourishing might reside outside the individual. It may be grounded in another person’s flourishing, or that of a group, institution, or society. Individuals are embedded within overlapping and interconnected layers of agency, such as groups and institutions. Questions of flourishing, languishing, and attentional normativity can arise for different levels, even if the relevant entity is not an agent. Individuals’ attentional demands might thus be grounded in traits or flourishing of some institution or group to which they belong.

Virtue theory can illuminate these interconnected levels of attentional demands. Conflicts can arise, for example, if attentional patterns serve the institution but stifle the individual. A virtue theory of attention can provide guidance on avoiding this, so that attentional interests better align. I preferred dishwashing in restaurants, rather than waiting tables, for example, because the cognitive monotony of dishwashing allowed me to become lost in thought. Others might prefer the higher attentional and cognitive demands and challenges of waiting tables. Understanding attentional virtues and flourishing might help evaluate working conditions by illuminating, for example, why employment in call centres is widely despised. Its attentional demands prevent the flow state without interpersonal or intellectual recompense, and attention is typically forced towards lousy topics. Fourthly, as intimated above, explaining the fundamental sources and grounds of normativity is perhaps not virtue theory's core aim.

A broader conception of the remit of virtue theory—encompassing virtue responsibilism, virtue ethics, and interwoven social layers of agency—can thus help illuminate various facets of the normativity of attention.

Acknowledgements

Many thanks to Adam Carter and Sandy Goldberg for their insightful and thought-provoking comments. I am grateful to Jon Garthoff for comments on an earlier draft and to Jordan Baker, Joseph Dartez, and Linh Mac for helpful discussions. This research was supported by an ACLS Fellowship from the American Council of Learned Societies.

Notes

- 1 Thanks to Jordan Baker and Jon Garthoff for helpful discussion.
- 2 Sosa (2017) *Epistemology* Princeton University Press.

3 From vice epistemology to critical character epistemology

Ian James Kidd

1 Introduction

A welcome development in recent epistemology has been the growing interest in epistemic vices, the negative character traits that stand opposed to what Linda Zagzebski named the virtues of the mind (Zagzebski 1996). Vice epistemology, named by Quassim Cassam, can be defined as ‘the philosophical study of the nature, identity, and epistemological significance of intellectual vices’ (Cassam 2016, 159). This recent interest in epistemic vices, which is a natural development of the earlier emergence of virtue epistemology in the early 1980s. A soberly honest stance on our personal and collective epistemic lives must acknowledge their susceptibilities to arrogance, dogmatism, closed-mindedness, and other failings of epistemic character. Without rushing into an optimistic stance on our capacities to overcome them, an important aspiration for vice epistemologists should be to try, as best we can, to find ways of minimising the incidence and severity of the vices of the mind—or, failing that, creating better ways of coping with their persistence within our lives.

I endorse the ameliorative spirit of vice epistemology, although in the absence of any definition of aims and success criteria, that may not be endorsing very much. There are very many things to which one can aspire concerning the epistemic vices, some more ambitious than others. At a minimum, we are coming to understand more about their nature, identity, and diversity and their ontological structures and relation to our human psychology. But we are also making some practical progress, too. Heather Battaly has excellent work on how we should modify features of our environments to mitigate our epistemically vicious tendencies (Battaly 2013, 2016). Alessandra Tanesini has excellent work showing how certain epistemic vices are constituted by stable psychological attitudes, which point to potential practical interventions (Tanesini 2016a, 2018). Further work with ameliorative potential continues to appear thanks to the current flow of interest in vice epistemology from epistemologists and those keen to put their work into practice.

The ameliorative potential of vice epistemology may depend, however, on certain methodological refinements. Much of how we ‘do’ vice

epistemology is informed by the tradition of Aristotelian virtue theory, which laid the basis of earlier work in virtue epistemology that, in turn, laid the basis for vice epistemology (Kotsonis forthcoming). Some vice theorists do draw upon other traditions, too, especially feminist and critical race epistemology. But there are reasons to think that Aristotelian approaches to epistemic vices lack some of the crucial sensitivities one needs to explore effectively certain aspects of character, virtue, and vice, of the sort, brilliantly articulated by Lisa Tessman (2005) and Robin Dillon (2012). But their work also points to potential reconstructions of Aristotelian character theories, some more radical than others. In what follows, I propose a reconstruction of vice epistemology, informed by Dillon's proposal for a *critical character theory*.

My aim is to present what, to honour Dillon's influence, I call a *critical character epistemology*. I sketch out its main features and show how it could, hopefully, better serve some of the ameliorative aims of those working to respond to epistemic vices. If it turns out those aims can be served without embracing a critical character epistemology, that's fine—we get the goods without needing the reforms. But it may also be that critical character epistemology has its own distinct merits. Before we can decide, though, we need to look more closely at the current state of vice epistemology.

2 Getting started in vice epistemology

We can find philosophical interest in arrogance, dogmatism, closed-mindedness, stupidity, indifference to the truth, and other epistemic vices among the earliest periods of the Greek, Indian, and Chinese traditions. Granted, their reasons for concern varied considerably, since their epistemological projects reflected their characteristic themes and concerns. Buddhist interest in epistemic vices, for instance, was tied into their fundamental soteriological aims. The story of the history of the philosophical study of epistemic character failings is not yet well understood, alas, though an impressive start has been made by historians of science and theology (DeYoung 2009; Kivisto 2014). Moreover, vice epistemologists, myself included, have tried to demonstrate the significance and interest in forms of historically informed vice epistemologies (Kidd 2021a). For instance, some of the earlier vice-epistemological projects had ameliorative aspirations, like the early modern English feminist vice epistemology we find in the work of Mary Astell and Mary Wollstonecraft (see Kidd 2018, §2A).

The earliest modern paper to use the term 'intellectual vice' (which I treat as synonymous with 'epistemic vice') was by Jonathan E. Adler (1999), who argued that while certain vices are harmful to enquiry, they are also vital for intellectual flourishing. Adler's paper was closely followed by Casey Swank's 2000 paper. Swank defined epistemic vices

as character traits ‘constitutive of unreasonableness’, which are therefore ‘bad in a specifically epistemic way’ (Swank 2000, 195). Unfortunately, those papers never got the uptake it merited. It was almost 15 years before widespread interest really got going. The main figure was Heather Battaly, who did three foundational things: she defended the legitimacy of agent-based appraisals from charges of *ad hominem*, she did the crucial conceptual work of distinguishing varieties of epistemic vice, and she provided a set of inspirational case studies of specific vices (see Battaly 2010, 2015). The latter included what I call *esoteric epistemic vices*—ones not currently entrenched in our inherited vice vocabularies, which helpfully expands our sense of the potential range of vices that ought to be on our investigative agenda. If we stick to the vices contingently present in our listings of the vices, we confine ourselves to a narrow, unscrutinised sense of the potential range of our epistemic character failings. Some esoteric epistemic vices include *epistemic insensibility* and *epistemic insouciance*, alongside other currently unnamed vices. For instance, Western theorising of the vices is deeply shaped by the concepts and concerns of Christian theology. We inherited rich concepts for theorising pride and other vices of humility, but are much less blessed when it comes to, for instance, the vices of curiosity (cf. Manson 2012; Pardue 2013).

It was common for virtue epistemologists to talk about vices, although usually only in passing, with the main business being exploration of epistemic excellences. An exception was Bob Roberts and W. Jay Wood, who offered ‘maps’ of various of the vices that gathered around the epistemic virtues they discussed (Roberts and Wood 2007). As Robin Dillon says, this may reflect the conviction that vices are ontologically and normatively secondary, that there is nothing to be gained by ‘looking at vice directly’ (Dillon 2012, 88). Robert Merrihew Adams, for one, argued that vices get less attention because ‘goodness is more fundamental than badness’ (Adams 2006, 36). Charlie Crerar names that conviction the *inversion thesis* and robustly rejects it. Roughly speaking, vices are not the ‘mirror images’ of virtues, because they have their own distinctive structures and features, which we are liable to miss if we simply create models of virtues and then invert them (Crerar 2018).

It is easy to encourage work on a topic when that work has a name and in the case of epistemic vices that baptism came with Quassim Cassam’s 2016 paper, ‘Vice Epistemology’. It came when there was a lot of that work to gather under that label. Battaly and Alessandra Tanesini had done a lot of work by then, of course, alongside the sustained analysis of the epistemic vices and injustices integral to systems of gendered and racial oppression offered by José Medina in his outstanding book *The Epistemology of Resistance*. He defines epistemic vices in terms of ‘a set of corrupted attitudes and dispositions’, which, if left unchecked, ensure that one’s ‘epistemic character tend[s] to become more corrupted’

(Medina 2012, 29, 72). Since then, there has been a burst of excellent work in vice epistemology including an edited collection in the *Journal of Philosophical Research* and the first dedicated monographs, Cassam's, *Vices of the Mind* (Cassam 2018) and Tanesini's *The Mismeasure of the Self* (Tanesini 2021).

The current work in vice epistemology is pleasingly pluralistic in its methods, aims, and inspirations. Aristotelian character theory, feminist epistemology, and intersectional social theory are drawn on alongside attitude psychology, critical race theory, and historical work documenting earlier ventures into the study of the vices of the mind. Much of the work also has an applied contemporary edge. Cassam's monograph, for instance, subtitled 'From the Intellectual to the Political', takes as its case studies recent political misadventures from Britain and the United States, from Brexit to the Trump Administration. In an age of flagrant public displays of vice, it may be no surprise that attention turns to vice theory.

Looking at current work in vice epistemology, there are three main sorts, although in practice they interpenetrate. To start with, there is *foundational work* on issues like the nature of epistemic vice, their relations to epistemic virtues and ethical vices, and the usual normative issues about how best to articulate their badness. The second sort of work are *case studies of specific vices*, detailed analyses of their structure, coupled to rich descriptions of their associated motivations, behaviours, and effects. Some of the well-studied vices include arrogance, dogmatism, closed-mindedness, hubris, insensibility, timidity, and servility.

The third sort of work is *applied vice epistemology*, putting these concepts to work in the effort to improve our epistemic conduct, practices, and systems. Roberts and Wood once referred to their work on epistemic virtues as a sort of regulative epistemology, a term they take from Nicholas Wolterstoffs (1996). A regulative epistemology, say Roberts and Wood, is one that seeks to 'generate guidance for epistemic practice', and is 'a response to perceived deficiencies in people's epistemic conduct, and thus is strongly practical and social' (Roberts and Wood 2007, 21). We could distinguish two types of regulative epistemology: one aimed at regulation of individual epistemic conduct, another aimed at active reform of our shared epistemic systems and practices. But that would be premature. Arguably the former cannot succeed without the latter given the complex ways that individual epistemic agency tends to be structured by our social environment—a point central to critical character epistemology and the wider traditions in feminist social philosophy to which it is indebted. I return to the collective dimensions of epistemic vices at the end of this chapter.

To summarise the points of this section, the study of epistemic character started in virtue epistemology during the 1980s, which dominated until the turn to epistemic vices in the last two decades. The focus on

epistemic virtues and flourishing is important and was a vital resource for vice epistemologists, for sure, although what is needed now is a corrective focus on the grimmer sides of epistemic life—on epistemic vices, failings, and corruption. To a degree, this has been helped along by the vigorous attention given these days to the many forms of epistemic violence (see, for instance, Berenstain 2016; Dotson 2016). This sort of perspective-broadening was advanced by Dillon: a critical character theorist aims ‘to understand moral character as affected by domination and subordination and by the struggles both to maintain and to resist and overthrow them’ (Dillon 2012, 84, 86).

From this perspective, we must change how we think about epistemic vices, too. The claim made by Dillon is that vices must be understood in terms of systems of domination and oppression and as characteristics of oppressors and as forms of damage done to those who are oppressed. A set of tight conceptual and causal connections obtained between vices and oppression must be acknowledged if progress is to be made in understanding and responding to either. If we look only at epistemic virtues and flourishing, then our vision of the world is not only *incomplete*—taking in only the brighter sides—but quite radically *distorting* in ways that occlude the realities of this world. It is the correction of this systematic distortion of epistemic character and agency that is the main aim of critical character epistemology.¹ The risk is that, without that darker, messier vision of human life, too many people will remain entrapped by the entrenched and ubiquitous patterns of vicious conduct that play out at the everyday and structural levels. Our lives must be understood, as Kate Norlock (2018) puts it, in the terms of a *perpetual struggle* focused on small, tangible acts of determined moral effort. On this view, any serious character ethics should accept that the ideal of flourishing is in reality the prerogative of the privileged. For the rest, what may be more realistic is the more existentially denuded aim of *coping* with the oppressive realities of the world.

Critical character epistemology is not pre-committed to anything as foreboding as the vision of perpetual struggle, although it should be honest about the sheer scale of the heavy ameliorative tasks that flow from its vision of the variety and tenacity of our many epistemic failings. It should also be clear why this is a *critical* character epistemology, since a key aim is scrutiny and revision, if necessary, of problematic epistemic conduct and the conditions that sustain it. Of course, there are other senses of criticism, too, like Kant’s, of establishing the conditions for the possibility of something.² Those may also apply, but that is not something I pursue in this chapter. Let’s now say more about epistemic vices and failings.

It should be clear, too, why it is a ‘critical’ character epistemology. Clear enough, at least, for me to move on to say more about epistemic vices and failings.

3 What are epistemic vices and why are they bad?

The complexity of our personal epistemic dispositions is the topic of study of character epistemology. For that reason, we should not think of that discipline as devolving into two relatedly independent enterprises, virtue epistemology, and vice epistemology. We need to study our excellences and failings of epistemic character together, rather than taking them in isolation then trying to weld the resulting accounts together. Since virtue epistemology is by now better developed and better known, I devote this section to surveying the current state of the art in vice epistemology. Along the way, I'll indicate why studying the vices of the mind cannot be done properly without constant reference to the virtues of the mind.

We can start with an ontological question, raised by Quassim Cassam (2020), which is: what kind of things are epistemic vices? Cassam argues the question devolves into three sub-questions: what *kinds* of things are epistemic vices, how do we *distinguish* different vices, and, to what are our distinctions between vices *answerable*? In response to the first question, there are two answers: a *vice-monist* says they are one kind of thing, a favourite answer being that they are *character traits*, an answer that goes back to Aristotle in the West. A *vice-pluralist*, however, allows that epistemic vices can be different kinds of things, including character traits, attitudes, and ways of thinking—or what Cassam neatly labels *character-vices*, *attitude-vices*, and *thinking-vices* (Cassam 2020, ch. 1). We see these kinds in vice epistemology. Battaly focuses on character-vices, Tanesini on attitude-vices, while a vice pluralism is endorsed by Cassam. Medina defines vices as ‘corrupted attitudes and dispositions’ and ‘attitudinal structures that permeate one’s entire cognitive life’ (Medina 2012, 30–31).

The second array of issues for vice epistemology is the set of normative questions about how best to understand the badness of epistemic vices, or, more specifically, to justify classification of a certain set of epistemic character traits, attitudes, or ways of thinking as *vices*. Sometimes it is clear *that* a certain epistemic character trait is bad, but less clear *what* is bad about it, and sometimes a fuller account of the badness of some trait only becomes clear once looked at using an appropriate normative framework. Within vice epistemology, there are two main normative models, each with its champions. *Vice-consequentialists* locate the badness of the epistemic vices in their effects and the best example is Cassam’s *obstructivism*, according to which epistemic vices are ultimately bad because they ‘systematically obstruct the gaining, keeping, and sharing of knowledge’ and other epistemic goods (Cassam 2018, 12). Battaly calls these *effects-vices* (Battaly 2014), which I divide into two sub-groups. *Productive effects-vices* are traits, like arrogance, that tend systematically to produce a preponderance of bad effects, while

passive effects-vices are traits, like epistemic laziness, that systematically fail to produce a preponderance of good effects. (Crudely, productive vices *do* bad, whereas passive vices *fail to do good*. In practice, of course, many vices do both, in which case we should just call them effects-vices.)³

The second normative model, best represented in the work of Alessandra Tanesini, is *vice-motivationalism*. It locates the badness of epistemic vices in the motivations, desires, and values of the epistemic agent. A vicious agent may be motivated by a desire to thwart the epistemic agency of others, or a desire to persist with beliefs that are comfortable even if also false, or the agent might value unearned confidence over humbling self-reflectiveness. Charlie Crerar usually distinguishes the two main types of vice-motivationalism (Crerar 2018, §§2–3). *Presence accounts* see vices as manifesting or revealing the *presence* of some epistemically bad motives, desire, or value, such as the desire to withhold salient information from other enquirers. *Absence accounts* see vices as manifesting the absence of some good motives, values, and desires, such as the lack of care or concern for truth, which is the heart of the vice Cassam calls *epistemic insouciance* (Cassam 2018, ch. 4). Jason Baehr, for one, has argued that ‘the most obvious or straightforward way a person can be intellectually vicious is motivational in nature: viz. by failing to care sufficiently about epistemic goods ... or by being outright opposed to them’ (Baehr 2020, 29).⁴

Alongside the consequentialist and motivationalist positions, there is, naturally, also a variety of pluralist positions. Such normative pluralism, as we might call it, can take several forms. One is that the badness of *all* epistemic vices can be articulated in consequentialist and motivationalist forms, with a proviso that, in some cases, references to effects won’t be enough. (I wonder, though, if this is a disguised form of motivationalism, since it relies on the claim that our analyses are deeper when they refer to motives). Another is that some of the vices can be satisfactorily appraised in consequentialist terms, others in motivationalist terms, and others still in more pluralist terms. I prefer that position, since pluralism of that sort seems a natural fit with the sheer variety and heterogeneity of our epistemic character failings. This latter sort of pluralism has a pragmatist streak: our question should be which of the available normative models do the job for any given epistemic vice, and we should not prejudge which model will be needed. Of course, when scrutinising that pluralism, we ought to attend to the familiar issues surrounding consequentialist and motivationalist normative theories—like the connection of intention to the outcome, the inscrutability of motives, and so on. At this point, there are rich prospects for more contacts between vice epistemology and normative ethics (see Baehr 2020, 33; Battaly 2014, ch. 4).

A critical character epistemologist is likely to embrace an ontological and normative pluralism about epistemic vices. Epistemic vices can be

many different kinds of things and they can be normatively appraised in reference to effects or inner states of the agent. This is consistent with their general pragmatism and desire to keep their options open, while also avoiding a bland sort of pluralism that says ‘everything goes’. But the ontological pluralism is perhaps quite radical. A character epistemologist, recall, takes as their focus excellences and failings of epistemic character, the main types of which are epistemic virtues and epistemic vices, respectively. But there are excellences of epistemic character that are not virtues and epistemic character failings that are not vices—at least, not on common conceptions of vice and virtue. Other excellences of epistemic character include a fantastic memory, a breadth and diversity of experience, various cognitive and perceptual skills, and a sense of maturity and degree of objectivity and reasonableness. I don’t think those are virtues, but they seem to be excellences of character adjacent to the epistemic virtues. Jason Baehr seems to share something like this view when he argues that intellectual virtues should be understood as ‘personal intellectual excellences’, as traits that ‘contribute to their possessor’s “personal intellectual worth”’ (Baehr 2011, 88–89). All virtues are excellences of character, but not all excellences of character are virtues.

A similar asymmetry holds for vices and failings of character. All vices are failings of character but not all failings of character are vices. Other epistemic failings include various cognitive biases, a narrowness and poverty of experience, a lack of crucial skills and abilities, and a lack of perspective and integrity (see, for instance, Holroyd 2020). Again, I don’t think those are vices in any familiar sense, but they are failings of epistemic character. Indeed, some of them are often defining *characteristics* of an epistemic agent, the sorts of features we might point out when giving an account of someone *qua* epistemic agent. A radical vice pluralist might just count them *all* as vices, but, for what it’s worth, that doesn’t sound right to me. Narrowness of experience is not a *vice*, even if it is sustained by vices, like arrogance.

Such issues about the definition of epistemic vices and failings might only exercise an enthusiastic vice epistemologist with ontological interests. If so, that’s fine. However we define the terms ‘vice’, ‘failing’, ‘excellence’, and ‘virtue’, we get the point that a character epistemologist is engaged in a careful, philosophical study of the nature, development, and significance of excellences and failings of epistemic character. Let’s now turn to two specific concepts central to their project.

4 From vices to predicaments

Epistemic vices have complex developmental histories. Many sources and conditions play a role in feeding their development and entrenchment within our epistemic character. A vice epistemologist is naturally

interested to explore those developmental processes, as complex as they will be. Robin Dillon emphasises that vices emerge and evolve through the complex interaction of psychological, interpersonal, developmental, and environmental processes or conditions. Character, therefore, should be conceived as ‘fluid, dynamic, and contextualised, both bodily and socially [and] as processive rather than substantive, as capable of stability without being static’ (Dillon 2012, 105). In an important remark for my present purposes, Dillon adds that

character dispositions [should] be understood to be inculcated, nurtured, directed, shaped, and given significance and moral valence as vice or virtue in certain ways in certain kinds of people by social interactions and social institutions and traditions that situate people differentially in power hierarchies.

(Dillon 2012, 104)

A critical character epistemologist inherits all of these insights and so searches for concepts that help us to articulate them. A vital concept is that of an *epistemic predicament*.

No one who lives in the social world could seriously think that it provides an Edenic environment that is maximally receptive to the cultivation and exercise of our epistemic capacities. The social world—or the variety of intermingled social worlds—is all messy and ridden with material, epistemic, and other suboptimalities. Some obvious examples include inequalities in distribution of goods, entrenched inequalities, problematic power relations, carefully maintained systems of collective ignorance, and entrenched systems of violence. Several generations of work by social epistemologists, feminist theorists, and activists have abundantly documented these and other suboptimalities (see, for instance, Bartky 1990; Collins 2000).

An obvious question is how issues of individual epistemic character relate to these wider social and structural conditions, since at first blush they may seem, methodologically at least, to proceed at very different levels. Dillon and other liberatory theorists emphasise, of course, that the situation is rather different—in her words, critical character theory (and epistemology) really ‘springs from the recognition that enslavement is not only social and material but also operates on and through character’ (Dillon 2012, 85). To develop this idea, we can turn to the concept of an epistemic predicament. It is developed in Medina’s book, though not systematically defined by him. He remarks, for instance, that our social identities and circumstances massively shape the sorts of concerns, dangers, needs, and risks that we are likely to experience—and, moreover, the sorts of resources and strategies available in our efforts to cope with those concerns. Medina, for instance, says that our predicaments affect whether and to what extent we labour under the burden of ‘lack of

access to information', 'lack of a credible voice and authority', persistent susceptibility to 'epistemic exclusions and injustices', and other predicamental challenges (Medina 2012, 29, 120).

Generalising from Medina's remarks, I will use the term 'epistemic predicament' to refer to the complex, contingent, and changing structure of epistemically-toned challenges, dangers, needs, and threats experienced by a person—an individual or a group—as a result of their particular emplacement within the social world. Three clarifications are needed for that definition. First, predicaments are radically *plural*, since they reflect the intersections of our multiple social identities. Ultimately, our predicaments might be unique, reflecting the subjectivity of each epistemic agent, even if the common structures of the social world tend to ensure a certain degree of commonality across the experiences of people sharing certain social identities.⁵ Second, predicaments are *ambivalent*—they cannot be neatly categorised as 'good' or 'bad', even if variable distributions of resources and opportunities favour certain predicaments in certain respects. Even highly privileged predicaments still incorporate *some* dangers and risks, even if these are lesser, qualitatively and quantitatively, than for those of oppressed social groups. Third, our epistemic predicaments are *changeable*, since they tend to reflect the stabilities and turbulence of the wider social world. People can try to change their epistemic predicaments in various ways, at the individual or collective level, and others can cooperate with or oppose those efforts. Conversely, one can also try to worsen the predicament of others by, for instance, subjecting them to epistemically violent behaviours (Dotson 2011).⁶

The concept of an epistemic predicament helps us to think in more socially sensitive ways about the development and perpetuation of epistemic vices and failings and therefore about the character-epistemic effects of social oppression. After all, it would be banal to say that 'human beings are prone to develop epistemic vices', since there are obvious variations and patterns in the prevalence of different vices across different groups of people. No doubt there are very complicated stories to tell about how different people acquire or develop the vices they do in the ways that they do. Medina, for instance, says that 'epistemic vices of all sorts are definitely possible outcomes of a socialisation under conditions of oppression', and emphasises that 'some epistemic vices are indeed more likely to be found among oppressed subjects' (Medina 2012, 40). His claim is not the obviously crude one that 'oppressed people develop ABC vices' and that 'privileged people develop XYZ vices': the subtler point is that 'the social positionality of agents does matter for the development of their epistemic character' (Medina 2012, 29, 40). Since that is a very general claim, we can add some more useful detail by appealing to the concept of epistemic predicament.

I propose that the particularities of our predicaments fundamentally structure the space of character-epistemic developmental possibilities a

person inhabits and also their ability to move through that space. There are many ways that can affect the relationship between vices and agents. Consider two: *susceptibility* and *salience*.

Starting with susceptibility, there is a very general sense in which all agents are to some degree susceptible to developing some or all vices. Anyone, in principle, could develop vices like arrogance, closed-mindedness, and mendaciousness. In practice, though, things will be more complex. There are often tangible patterns of susceptibility, shaped by subjective, social, and structural factors as well as, in some cases, bad epistemic luck (although see Berenstain (forthcoming) for salutary warnings about attributing to bad epistemic luck processes that in fact are part of systems of oppression). To take an example, those with multiply privileged identities may be more systematically susceptible to the ego-inflationary epistemic vices like arrogance and haughtiness (cf. Tanesini 2016b; 2018). As Medina emphasises, belonging to a privileged group is neither necessary nor sufficient for the development of epistemic vices (Medina 2012, 40). Many actions and contingencies can intervene to realise or suppress the susceptibilities that confront us in our efforts to navigate the vice-conducive pressures and temptations of the social world. For that reason, one very important protective capacity will be what Medina calls ‘lucidity’ about our epistemic predicament—at a minimum, a sense of which vices or clusters of vices lie in one’s path as upcoming or tangible risks, and which, by contrast, safely lie well outside one’s path.⁷

A second way that predicaments can shape our character-epistemic developmental possibilities for the worst concerns the *salience* of different epistemic vices. In a general sense, all epistemic vices are salient to some degree, since all of them will stand out as significant in some sense: a vice may appear as alarming, horrifying, irritating, serious, trivial, and so on. I expect most people would regard, say, arrogance and manipulateness as worse vices than, say, incuriosity and superficiality. The salience of epistemic vices depends on many different factors, many of which are refracted through our specific predicaments. A good example is the fact that members of some social groups are negatively stereotyped as being *essentially* prone to or *characterised by* certain vices—women, for instance, as banal, incurious, unreflective, and so on. Mary Astell wrote in 1694 of the entrenched expectations of her society that women, by virtue of their ‘degraded reason’, necessarily suffered from a ‘degenerated and corrupted’ epistemic character, incapable of sustaining epistemic virtues. Astell was alert to the culturally reinforced expectation that women were, or would always become, marked by the ‘Feminine Vices’, like submissiveness and superficiality. Within that misogynistic social and epistemic culture, gendered vices become especially salient to women who reflect on their characters or seek to improve their epistemic predicament (Astell 2002, 62).

A critical character epistemologist can use the concept of epistemic predicaments to think about epistemic character and vices in relation to the specifics of our emplacement in the social world. By thinking in terms of predicaments, we can go beyond abstract accounts, and talk in more discerning ways about the ways that our susceptibility to specific epistemic vices, and the specific salience of those vices, is shaped by our predicaments. Naturally, the task is complicated. Epistemic predicaments are plural, changing, and intersectional; some vices are highly gendered and racialised and some are embedded in wider cultural or moral conceptions. But this is the price we pay for the sorts of social-sensitive study of epistemic vices that we need to ensure we are tracking the complex connections between epistemic agents and social structures.

5 From predicaments to corruption

A critical character epistemologist wants to explore the specific patterns of susceptibility to epistemic vices for differently situated groups of epistemic agents. Thinking in terms of the predicaments people face can help with that task. But thinking in terms of susceptibilities and of salience only tells us about which vices we might develop, and which might stand out for us. It also shapes what sorts of virtues—or what specific inflections of the virtues—are pertinent to our self-protective strategies (Monypenny 2021). We also need to ask *how* people actually acquire the vices to which they are susceptible and which they presumably *want* to avoid, given the negative salience of those vices. (I assume it is more important to try to avoid developing a vice that is judged to be more alarming or worrisome.) To answer that question, we need to add the concept of *epistemic corruption*.

A vocabulary of corruption often features within vice-theoretic discourses. Gabriele Taylor remarks that moral ‘vices corrupt and destroy’ (Taylor 2006, 126) while Judith Shklar remarks that vices ‘dominate and corrupt’ our character (Shklar 1984, 200). We also find the language of corruption in vice epistemology. For Miranda Fricker, internalisation of sexist and racist norms, values, and assumptions ‘corrupts’ our epistemic sensibilities and in that way can ‘inhibit’ and ‘thwart’ the development of epistemically virtuous character (Fricker 2007, 59, 58, 30). José Medina, recall, defines epistemic vices in terms of ‘corrupted attitudes and dispositions’, and argues that, under oppressive conditions, one’s ‘epistemic character [will] tend to become more corrupted’ (Medina 2012, 29, 72).

Although none of these writers used the term ‘corrupt’ in a technical sense, they use it to refer to a phenomenon specific to critical character epistemology. One of the main ways that agents become epistemically vicious is that they are subjected to processes and conditions that are *epistemically corrupting*—a concept I have developed elsewhere (see Kidd 2019, 2020). On my account, epistemic corruption occurs when one’s

epistemic character comes to be damaged due to one's interaction with *corruptors*—conditions, processes, doctrines, or social structures that tend to facilitate the development and exercise of epistemic vices. Corruption is dynamic and also diachronic, typically consisting of sustained exposure to corruptors, rather than singular events. The term 'facilitate' includes 'encourage', 'promote', 'incentivise', and 'provides inducements, rewards, and temptations to acts of epistemic vice'. There are several modes of corruption, of which the main ones are:

- 1 *Acquisition* of novel epistemically vicious attitudes, character traits, and ways of thinking, of a sort not previously a feature of the subject's epistemic character.
- 2 *Activation* of epistemically vicious attitudes, character traits, and ways of thinking that are present in the subject's epistemic character but dormant and inactive.⁸

The next three modes are different: they involve amplification of certain aspects of whichever epistemic vices are already active:

- 3 *Propagation* occurs when corrupting conditions increase the scope of a vice, viz., the extent to which it affects one's epistemic activities. In Annette Baier's useful remark, an initially localised vice propagates when it starts to 'infect their whole character' (Baier 1995, 274).
- 4 *Stabilisation* occurs when corrupting conditions increase the *stability* of a vice. Some vices are unstable, flickering 'on and off', under the positive counteracting influence of acts of willpower, social censure, or whatever. As vices stabilise, though, they become more resistant to destabilisation.
- 5 *Intensification* occurs when corrupting conditions increase the *strength* of a vice. The vices in their weaker forms tend to produce fewer bad effects and express weaker bad motives. But vices can be strengthened, making them more intense and extreme and therefore become ever-more problematic.

The social world is filled with potential corruptors that can act on our epistemic characters by facilitating our complex predicamental susceptibilities to epistemically vicious attitudes, character traits, and ways of thinking. A critical character epistemologist will be very keen to study the conceptual and causal relationships between vices, corruptors, and characters (cf. Battaly 2013; Cooper 2008).⁹ This calls for integrated vice-epistemological and empirical research of the sort already profitably taken by moral psychologists interested in the virtues (see, e.g., Miller 2017; Snow 2014).

I think that the social world is vastly epistemically corrupting and that our epistemic predicaments structure the diversity and intensity of the epistemically corrupting influences that we have to navigate. That

includes the vices that are salient to me and to which I'm susceptible and the specific types of corruptors that loom large in my social experiences, not to mention the sorts of counter-corrupting influences and resources on which I can try to draw in order to protect the fragile mesh of virtuous dispositions that make up the better parts of my epistemic character. A universal feature of all epistemic predicaments is the task of trying to avoid or manage those corrupting influences and structures while trying to simultaneously minimise the character damage one suffers and also trying to fulfil the many other pressing demands of one's epistemic and social life. Struggling against the perpetually present risks of epistemic corruption is only ever a part of the business of trying to live well.

A key task of critical character epistemology is to develop a working understanding of the variety of corruptors out there in the world, partly to guide the empirical research but also as a way of training our epistemic sensibilities. To that end, consider some general sorts of corruptors that the critical character epistemologist wants to identify and, ultimately, try to either remove, reform, or avoid:

- 1 The absence or derogation of epistemic exemplars or 'heroes', who practically model forms of epistemic virtue, excellence, and integrity (see Croce and Vaccarezza 2017; Zagzebski 2017).
- 2 The valorisation and elevation of exemplars of epistemically vicious persons and acts by, for instance, ensuring that they receive social goods such as authority, respect, and power.
- 3 The rebranding of vices as virtues in ways that can prevent someone from detecting that they are being corrupted (see Dillon 2012, 99). Sometimes, a person might be genuinely unaware they are becoming corrupted, not least given that certain vices have a self-concealing capacity—so-called *stealthy vices* (Cassam 2018, ch. 7).
- 4 The establishment of conditions that increase the exercise costs of virtues. One can make it harder to exercise certain epistemic virtues by, say, depriving a person of the necessary amounts of *time* or reacting to acts of epistemic courage with an elevated threat of violence (Kidd 2022).¹⁰
- 5 The establishment of conditions that increase the incentives to vice. By arranging an environment to incentivise and reward acts of vice, one can habituate people to acts of vice that, over time, can transform their epistemic character for the worse.

These are some of the main types of corruptors, described very generally, each inviting more investigation. Alongside their general relevance to vice epistemology, they are of particular significance to a critical character epistemologist. Many of those corruptors are themselves implicated in wider systems of oppression. José Medina, for instance, describes *epistemic heroes*, 'extraordinary subjects who under conditions

of epistemic oppression are able to develop epistemic virtues with a tremendous transformative potential' (Medina 2012, 186). Obviously, such epistemic heroes are often characterised by the virtue of epistemic courage, and a natural response of oppressors to such heroes is to derogate and assail them—a clear case where an oppressive system tries, often successfully, to massively increase the exercise costs of epistemic virtues (see, further, Kidd 2018).

The deep relationship between processes of epistemic corruption and oppressive social systems is one reason why the ameliorative goals of a critical character epistemology necessarily take on an overtly political character. When characterising the ultimate aims of critical character theory, Dillon quotes Max Horkheimer's explanation that the aim of critical theory is 'to liberate human beings from the circumstances that enslave them' (quoted in Dillon 2012, 85). Systems of enslavement act on and through character, including through a complex web of epistemically corrupting processes and structures that damage and distort the epistemic character of subjects, the oppressors and the oppressed alike.¹¹ It is in relation to that socially transformative goal that critical character epistemology should ultimately be understood.

To summarise: current work in vice epistemology offers powerful ways of thinking in systematic detail about the variety of failings of epistemic character to which human beings are susceptible. Such susceptibilities arise from our psychological and cognitive limitations, the abrasive effects of so many of our interpersonal encounters and relationships, the suboptimalities of our social worlds, and the systems of oppression characteristic of so many of those worlds. I have described a specific style of vice epistemology—*critical character epistemology*—and some of its distinguishing features. These include its adoption of the concepts of epistemic predicaments and epistemic corruption and the explicit socio-political goals that align it in many ways with wider progressive social movements. I do not think that all of those with an interest in epistemic vices need to be critical character epistemologists. But I do think that a vice epistemologist with liberatory aspirations might find critical character epistemology an ally in their efforts.

Acknowledgements

I am grateful for the discussion and encouragement with the attendees of the COGITO Epistemic Vices workshop, hosted by the University of Glasgow.

Notes

¹ It is interesting to notice that although we have a well-developed tradition in virtue theory, there is hardly anything we could call *vice ethics*. Granted,

- there are honourable exceptions, like Lisa Tessman (2005). There are also those who urge relevantly grim estimations of our collective moral and epistemic condition, like David E. Cooper (2018) and Kate Norlock (2018). Indeed, I argue elsewhere that our many failings are so diverse, entrenched and ubiquitous that they justify a charge of misanthropy, a critical verdict on our collective moral condition (Kidd 2021b).
- 2 I thank Mark Alfano for this useful point about the different senses of ‘criticism’.
 - 3 A critical amendment to obstructivism is offered by Kotsonis (2022).
 - 4 Crerar also adds a third ‘compatibility’ position, which sees some vices, at least, as being composed of intermingled virtuous and vicious motivations: think of a conspiracy theorist who is radically doxastically rigid, but also genuinely driven by a conscientious commitment to the truth.
 - 5 Medina speaks of the predicaments of the privileged and of the oppressed, although would likely emphasise their heterogeneity (Medina 2012, §§ 1.1–1.2).
 - 6 The term ‘epistemic violence’ was introduced by Gyatri Spivak (1998).
 - 7 I am thinking here of Wittgenstein’s remark: ‘[t]here are problems I never tackle, which do not lie in my path or belong to my world’ (Wittgenstein 1998, 11).
 - 8 A vice-consequentialist might not recognise the existence of dormant traits, since they are not producing any bad epistemic effects. But dormant vices would, if activated, produce bad effects, so vice-consequentialist should still worry about them.
 - 9 An important distinction to consider is that between *monocorrupting* and *polycorrupting* conditions: those that facilitate one single vice and those that facilitate a broader range of vices. Is it the case, for instance, that an epistemically homogeneous environment can corrupt for a whole range of vices?
 - 10 Consider, for instance, the procedural epistemic virtues, like carefulness, diligence, and thoroughness (Kidd 2022).
 - 11 Compare Lisa Tessman on the two types of ‘moral damage’—that is, damage to the moral character of people—integral to systems of oppression (Tessman 2005, chs. 2 and 3).

References

- Adams, Robert Merrihew (2006) *A Theory of Virtue: Excellence in Being for the Good* (Oxford: Oxford University Press).
- Adler, Jonathan E. (1999) “Epistemic Dependence, Diversity of Ideas, and a Value of Intellectual Vices”, *The Proceedings of the Twentieth World Congress of Philosophy* 3: 117–129.
- Astell, Mary (2002) *A Serious Proposal to the Ladies. Parts I and II* [1694]. Ed. Patricia Springborg (Ontario: Broadview Literary Texts).
- Baehr, Jason (2011) *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology* (Oxford: Oxford University Press).
- Baehr, Jason (2020) “The Structure of Intellectual Vices”, in Ian James Kidd, Heather Battaly and Quassim Cassam (eds.), *Vice Epistemology* (New York: Routledge), 21–36.
- Bajer, Anette (1995) *Moral Prejudices: Essays on Ethics* (Cambridge, MA: Harvard University Press).
- Bartky, Sandra-Lee (1990) *Femininity and Domination: Studies in the Phenomenology of Oppression* (New York: Routledge).

- Battaly, Heather (2010) "Attacking Character: *Ad Hominem* Argument and Virtue Epistemology", *Informal Logic* 30(4): 361–390.
- Battaly, Heather (2013) "Detecting Epistemic Vice in Higher Education Policy: Epistemic Insensibility in the Seven Solutions and the REF", *Journal of Philosophy of Education* 47 (2): 263–280.
- Battaly, Heather (2014) "Varieties of Epistemic Vice", in Jon Matheson and Rico Vitz (eds.), *The Ethics of Belief*. (Oxford: Oxford University Press), 51–76.
- Battaly, Heather (2016) "Developing Virtue and Rehabilitating Vice: Worries about Self-Cultivation and Self-Reform", *Journal of Moral Education* 45: 207–222.
- Battaly, Heather (2019) "Vice Epistemology Has Aa Responsibility Problem", *Philosophical Issues* 29 (1): 24–36.
- Berenstein, Nora (2016) "Epistemic Exploitation", *Ergo* 3: 569–590.
- Berenstein, Nora (2020) "White Feminist Gaslighting", *Hypatia* 35(4):733–758.
- Cassam, Quassim (2016) "Vice Epistemology", *The Monist* 99 (3): 159–180.
- Cassam, Quassim (2018) *Vices of the Mind: From the Intellectual to the Political* (Oxford: Oxford University Press).
- Cassam, Quassim (2019) *Vices of the Mind: From the Individual to the Political* (Oxford: Oxford University Press).
- Cassam, Quassim (2020) "The Metaphysics of Epistemic Vice", in Ian James Kidd, Heather Battaly and Quassim Cassam (eds.), *Vice Epistemology* (New York: Routledge), 37–52.
- Collins, Patricia Hill (2000). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*, revised ed. (New York: Routledge).
- Cooper, David E. (2008) "Truthfulness and Teaching", *Studies in Philosophy and Education* 27: 79–87.
- Cooper, David E. (2018) *Animals and Misanthropy* (London: Routledge).
- Crerar, Charlie (2018) "Motivational Approaches to Intellectual Vice", *Australasian Journal of Philosophy* 96(4): 753–766.
- Croce, Michel and Maria Silvia Vaccarezza. (2017) "Educating Through Exemplars: Alternative Paths to Virtue", *Theory and Research in Education* 15 (1): 5–19.
- DeYoung, Rebecca Konyndyk (2009) *Glittering Vices: A New Look at the Seven Deadly Sins and Their Remedies* (Grand Rapids, MI: Brazos Press).
- Dillon, Robin (2012) "Critical Character Theory: Toward a Feminist Perspective on 'Vice' (and 'Virtue')", in Sheila L. Crasnow and Anita M. Superson (eds.), *Out from the Shadows: Analytical Feminist Contributions to Traditional Philosophy*, (New York: Oxford University Press), 83–114.
- Dotson, Kristie (2011) "Tracking Epistemic Violence, Tracking Practices of Silencing", *Hypatia* 26 (2): 236–257.
- Dotson, Kristie (2016) "Contextualising Epistemic Oppression", *Social Epistemology* 28(2): 115–138.
- Fricker, Miranda (2007) *Epistemic Injustice: Power and the Ethics of Knowing* (Oxford: Oxford University Press).
- Holroyd, Jules (2020) "Implicit Bias and Epistemic Vice", in Ian James Kidd, Heather Battaly, and Quassim Cassam (eds.), *Vice Epistemology* (New York: Routledge), 126–147.
- Kidd, Ian James (2016) "Charging Others with Epistemic Vice", *The Monist* 99(3): 181–197.

- Kidd, Ian James (2018) "Epistemic Courage and the Harms of Epistemic Life", in Heather Battaly (ed.), *The Routledge Handbook to Virtue Epistemology* (New York: Routledge), 244–255.
- Kidd, Ian James (2019) "Epistemic Corruption and Education", *Episteme* 16(2): 220–235.
- Kidd, Ian James (2020) "Epistemic Corruption and Social Oppression", in Ian James Kidd, Heather Battaly and Quassim Cassam (eds.), *Vice Epistemology* (New York: Routledge, 2020), 69–86.
- Kidd, Ian James (2021a) "A Case for an Historical Vice Epistemology", *Humana.Mente* 14(39): 69–86.
- Kidd, Ian James (2021b) "Varieties of Philosophical Misanthropy", *Journal of Philosophical Research* 46: 27–44.
- Kidd, Ian James (2022) "Character, Corruption, and 'Cultures of Speed' in the Academy", in Áine Mahon (ed.), *Philosophical Perspectives on the Contemporary University: In Shadows and Light* (Dordrecht: Springer), 17–28.
- Kivisto, Sari (2014) *The Vices of Learning: Morality and Knowledge in Early Modern Universities* (Leiden: Brill).
- Kotsonis, Alkis (2022) "A Novel Understanding of the Nature of Epistemic Vice", *Synthese* 200(1): 1–16.
- Kotsonis, Alkis (forthcoming) "The Aristotelian Understanding of Intellectual Vice: Its Significance for Contemporary Vice Epistemology", *European Journal of Philosophy*.
- Manson, Neil C. (2012) "Epistemic Restraint and the Vice of Curiosity", *Philosophy* 87: 239–259.
- Medina, José (2012) *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations* (Oxford: Oxford University Press).
- Midgley, Mary (1984) *Wickedness: A Philosophical Essay* (London: Routledge and Kegan Paul).
- Miller, Christian (2017) *The Character Gap: How Good Are We?* (Oxford: Oxford University Press).
- Monypenny, Alice (2021) "Between Vulnerability and Resilience: A Contextualist Picture of Protective Epistemic Character Traits", *Journal of Philosophy of Education Society of Great Britain* 55 (2): 358–370.
- Norlock, Kathryn (2018) "Perpetual Struggle", *Hypatia* 34(1): 6–19.
- Pardue, Stephen T. (2013) *The Mind of Christ: Humility and the Intellect in the Early Christian Tradition* (London: Bloomsbury).
- Roberts, Robert C. and W. Jay Wood (2007) *Intellectual Virtues: An Essay in Regulative Epistemology* (Oxford: Oxford University Press).
- Shklar, Judith (1984) *Ordinary Vices* (Cambridge, MA: Harvard University Press).
- Snow, Nancy, ed. (2014) *Cultivating Virtue: Perspectives from Philosophy, Theology, and Psychology* (Oxford: Oxford University Press).
- Spivak, Gayatri (1998) "Can the Subaltern Speak?", in Cary Nelson and Lawrence Grossberg (eds.), *Marxism and the Interpretation of Culture* (Urbana: University of Illinois Press), 271–313.
- Swank, Casey (2000) "Epistemic Vice", in Guy Axtell (ed.), *Knowledge, Belief, and Character: Readings in Virtue Epistemology* (New York: Rowman & Littlefield Publishers), 195–204.

- Tanesini, Alessandra (2016a) "Teaching Virtue: Changing Attitudes", *Logos and Episteme* 7(4): 503–527.
- Tanesini, Alessandra (2016b) "'Calm Down, Dear': Intellectual Arrogance, Silencing and Ignorance", *Aristotelian Society Supplementary Volume* 90(1): 71–92.
- Tanesini, Alessandra (2018) "Intellectual Servility and Timidity", *Journal of Philosophical Research* 43: 21–41.
- Tanesini, Alessandra (2021) *The Mismeasure of the Self: A Study in Vice Epistemology* (Oxford University Press).
- Taylor, Gabriele (2006) *Deadly Vices* (Oxford: Clarendon).
- Tessman, Lisa (2005) *Burdened Virtues: Virtue Ethics for Liberatory Struggles* (Oxford University Press).
- Wittgenstein, Ludwig (1998) *Culture and Value: A Selection from the Posthumous Remains*, edited by G. H. von Wright in collaboration with Heikki Nyman, revised of the text by Alois Pichler, translated by Peter Winch (Oxford: Blackwell).
- Wolterstorff, Nicholas (1996) *John Locke and the Ethics of Belief* (Cambridge: Cambridge University Press).
- Zagzebski, Linda (1996) *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge* (Cambridge: Cambridge University Press).
- Zagzebski, Linda (2017) *Exemplarist Moral Theory* (Oxford: Oxford University Press).

3b Commentary from Heather Battaly

Comments on Ian James Kidd's 'From Vice Epistemology to Critical Character Epistemology'

HEATHER BATTALY

Ian James Kidd's chapter argues that feminist character theory has important insights for vice epistemology. One of those insights is for vice epistemology's ameliorative wing, which explores strategies for reforming epistemic vices, and (relatedly) the causes and etiologies of epistemic vices. Kidd draws inspiration from feminist analyses of character, especially the critical character theory pioneered by Robin Dillon (2012). In so doing, he proposes a critical character epistemology that recognises the influence of oppressive social structures on the development of epistemic vices in individuals. Specifically, he contends that social structures and conditions can be 'corrupting', that is, they can promote, encourage, and incentivise epistemic vices in individuals. Moreover, they can corrupt different individuals in different ways, facilitating (e.g.) intellectual arrogance in one person, and intellectual servility in another. Accordingly, a key insight of Kidd's critical character epistemology is that we won't be able to reform epistemic vices in individuals without also reforming the oppressive social structures that facilitate them. As he puts this point elsewhere, relying on strategies that target changes in individuals without addressing the reform of corrupting social structures would be 'a febrile form of ameliorative whack-a-mole' (Kidd 2020: 80). I think Kidd's argument is doing laudable and crucial work at the intersection of vice epistemology, liberatory epistemology, and feminist character theory. It has the added bonus of making a number of other helpful points along the way—for example, Kidd distinguishes between productive and passive effects-vices, suggests that epistemic vices are widespread in the real world due to oppressive social structures (they aren't solely possessed by high-profile political figures), and identifies several types of 'corruptors' and corrupting conditions including the valorisation of epistemically vicious persons. Below I ask three sets of questions about the implications of Kidd's argument.

First, Kidd argues that ‘one of the main ways that agents become epistemically vicious is that they are subjected to corrupting processes and conditions’. In other words, people can become vicious—they can ‘actually acquire the vices to which they are susceptible’—by interacting with corruptors. This leads to a set of questions about whether individuals can be blameworthy for their epistemic vices. Does critical character epistemology allow for this? Does it allow individuals to be blameworthy in the sense that they are accountable? Perhaps it doesn’t: if individuals can become epistemically vicious through their interactions with corruptors, then they may not exercise enough control over the acquisition of their vices to be accountable for coming to have them. But, perhaps it does: if corruption and the acquisition of vice aren’t inevitable, and if individuals can recognise corruptors for what they are and sometimes avoid interacting with them, for example, by working with allies to construct islands of ‘epistemic edification’ (Kidd 2019), then they may exercise enough control to be accountable. Even if critical character epistemology doesn’t allow individuals to be accountable for their epistemic vices, could it endorse a sort of blameworthiness that doesn’t entail control? Might individuals be blameworthy for their epistemic vices in the sense that their vices are attributable to them, or in the sense that they are answerable for them, or in the sense that they are reprehensible for them (Cassam 2019; Tanesini 2021)? Relatedly, does critical character epistemology assign a role to forward-looking responsibility—to individuals *taking* responsibility for their vices? Does it assign a role to charging others with epistemic vices (Kidd 2016)? To be sure, this is a wide-ranging set of questions, which cannot be answered quickly! My hope is that Kidd can point out some routes that are open to critical character epistemology, giving us some promising directions to explore.

Second, Kidd’s chapter emphasises the role that social structures play in facilitating epistemic vices in individuals. This is one important way in which epistemic virtues and vices can be social, namely, their development can be social. As Kidd suggests, if social structures are oppressive and corrupting, epistemic vices may even be endemic. This is a point he spotlights, perhaps because it is sometimes overlooked. I’d be interested to hear Kidd’s thoughts about another way in which epistemic vices might be social. Can oppressive social structures and institutions themselves have epistemic vices, that is, can ‘corruptors’ have vices? Must social structures have epistemic vices in order to be corrupting? Or, can they corrupt (facilitate vices) without having any vices themselves? Presumably, a corrupting structure need not possess a particular vice in order to facilitate it—arguably, structures that are intellectually arrogant can facilitate intellectual servility in some individuals who interact with them. But, must corruptors have some vice or other? Relatedly, can corruptors have some epistemic vices that they don’t facilitate in any individuals? More broadly, what are the implications of structural vices

for competing theories of the nature of epistemic vice? Are structural vices easier for obstructivist accounts (Cassam 2019) to accommodate, or can motivationalist accounts (Tanesini 2021) do an equally good or better job?

Finally, I close with a set of questions about potential next steps. Kidd argues that we will need to reform corrupting structures if we hope to be effective in reforming epistemic vices in individuals. Presumably, structural reform will be slow and involve solidarity. I'd welcome Kidd's ideas about where and how to begin, and whether critical character epistemology can suggest some potential strategies. Perhaps, we can try to reverse the conditions of corruption that Kidd identifies, that is, reverse the derogation of virtuous exemplars, the valorisation of vicious exemplars, and incentives to vice. If corrupting structures themselves have vices, we may also need to reform those. Can we make progress in reforming corrupting structures by facilitating liberatory epistemic virtues, such as meta-lucidity, in individuals (Medina 2012)? Will we also need to facilitate liberatory epistemic virtues in structures themselves? If epistemic vices are stealthy, then can we expect strategies for facilitating epistemic virtues to be effective in reforming epistemic vices? Or will we need some different strategies for reforming epistemic vices?

References

- Cassam, Quassim. 2019. *Vices of the Mind*. Oxford: Oxford University Press.
- Dillon, Robin. 2012. "Critical Character Theory: Toward a Feminist Perspective on 'Vice' (and 'Virtue')." In Sharon L. Crasnow and Anita M. Superson (eds.) *Out from the Shadows: Analytical Feminist Contributions to Traditional Philosophy*. Oxford: Oxford University Press, 83–114.
- Kidd, Ian James. 2016. "Charging Others with Epistemic Vice." *The Monist* 99(2): 181–197.
- Kidd, Ian James. 2019. "Epistemic Corruption and Education." *Episteme* 16(2): 220–235.
- Kidd, Ian James. 2020. "Epistemic Corruption and Social Oppression." In Ian James Kidd, Heather Battaly and Quassim Cassam (eds.) *Vice Epistemology: Theory and Practice*. New York: Routledge, 69–87.
- Medina, José. 2012. *The Epistemology of Resistance*. Oxford: Oxford University Press.
- Tanesini, Alessandra. 2021. *The Mismeasure of the Self*. Oxford: Oxford University Press.

3c Commentary from S. Goldberg

The Banality of Vice

GEORGI GARDINER

Kidd argues that vice epistemology is fruitfully developed as critical character epistemology.¹ He outlines three hallmarks of critical character epistemology, which it shares with forebears such as critical race theory, feminist epistemology, and—more directly—Robin Dillon’s critical character theory. The first hallmark is social critique. Critical character epistemology theorises harms and injustices, focusing on systems of domination and subordination. Kidd argues that current epistemology disproportionately focuses on epistemic goods, such as virtue. Foregrounding aphotic and unpropitious facets of social-epistemic life, including vice, is a needed corrective. Secondly, critical character epistemology aims to ameliorate current conditions. Thirdly, it highlights interconnections between the individual and their society, especially how social forces shape epistemic character. This is the epistemic analogue of Dillon’s (2012, 85) claim that ‘enslavement is not only social and material but also operates on and through character’.

Social position, such as race and class, affects which character traits are differentially nurtured and discouraged, which benefit or impede us, and to which vices we are particularly susceptible. Kidd investigates these relations between social position and epistemic character development. I focus on the effects of salience distributions—specifically the relative salience of vices—on how social position affects epistemic character.

Kidd claims ‘all epistemic vices are salient to some degree, since all of them will stand out to us as significant in some sense—vice may appear as alarming, horrifying, irritating, serious, trivial, and so on’. He notes the salience of particular vices can depend on social position. Kidd writes, ‘A good example is the fact that members of some social groups are negatively stereotyped as being *essentially* prone to or *characterised* by certain vices—women, for instance, as banal, incurious, unreflective,

and so on' (emphasis in original). This stereotyping affects the salience of vices. Drawing on the 1694 writing of Mary Astell, Kidd writes,

Astell was alert to the culturally reinforced expectation that women were, or would always become, marked by the 'Feminine Vices', like submissiveness and superficiality. Within that misogynistic social and epistemic culture, those gendered vices become especially salient to women seeking to improve their epistemic predicament.

Kidd is correct that salience plays crucial roles in how social position affects epistemic character development. But I don't think Kidd aptly sketches these roles. I sketch alternative ways the differential salience of vice influences character development.

Firstly, I disagree with Kidd's contention that 'all epistemic vices are salient'. Indeed, a critical character epistemologist should take particular issue with this claim. Salience is the property of being attention-grabbing; it reflects descriptive, rather than prescriptive, facts. Salient things have cognitive prominence. We must distinguish this from what is important, relevant, or concomitant. Some moral facts might be important, for example, but are typically overlooked and so not salient.

Social inequalities, including in distributions of epistemic traits and expectations about those traits, can be more pernicious when overlooked. The epistemic vices of chauvinism, white ignorance, and unquestioning deference to one's birth culture and religion, for instance, are widespread in part because they are not grokked. They compose part of the background tapestry of society. The vices of bias are often unnoticed, rendering them harder to correct. Similarly, we expect wealthy people to exhibit high confidence in their beliefs and abilities, making it difficult recognise vicious overconfidence.

Epistemic vices might be more salient to those who suffer their effects. Women are more apt to identify sexism, for example. But, firstly, current attunement to such vices benefits from decades of feminist theorising; prior to this, chauvinist epistemic vices would often be inaccurately viewed as the person's having apposite beliefs. Secondly, even those injured by the vice might not see it as vice. A daughter might suffer consequences of having sexist parents, for example, yet not recognise their traits as sexist epistemic vices.

Vice can be like air—invisible, unnoticed, camouflaged by ubiquity. Charles Mills highlighted the pervasion of white ignorance; critical character epistemology should emphasise how vice can be similarly banal. Vice is the normal condition of everyday lives.

Men's emotions can enjoy a similar inconspicuousness. Society shapes it, and individuals contort around it, without fully appreciating

its presence or seeing it *as* emotion. Part of the social potency of some emotions, vices, and virtues stems from their being rendered invisible or mistaken for something else, such as pure ‘rationality’.

Secondly, specific stereotyped epistemic traits have a complex relationship to salience. I’ll use Kidd’s seventeenth-century example of stereotypes of women as incurious. Given this, women are expected to be incurious. The expectation constitutes and reinforces the norm. People might not notice the expectation unless it is violated. Incuriousness in women is not remarked upon; instead, curiosity is salient. Departures from normed vices, or attempts to shed normed vices, attract attention. Deviance is noticed. And this salience, with its concomitant censure or risk, disincentivises the aberrance. This helps explain why people conform; it can be safer to not stand out.

For some groups, character traits are noticed but the perceived valence is switched or downplayed. One might notice the elevated confidence of wealthy people, for example, but not perceive it as bad. It might instead garner respect, emulation, or deference. Or one might see it as bad but tend to downplay or overlook it. To illustrate the distinct role of attention, suppose pop musicians are stereotyped as incurious and gender-nonconforming youths as epistemically mercurial; they are seen as changing their minds frequently. A person might regard each property as—let’s say—equally bad. But only the latter and the latter’s badness are salient to him. When he thinks of celebrities, he seldom remembers their incuriosity. When he thinks of gender-non-conforming youths, by contrast, he often remembers their perceived epistemic caprice. The social privilege of celebrities, and relative marginalisation of trans youths, bolsters—and partly comprises—these attention patterns.

Kidd suggests that ‘gendered vices [like submissiveness and superficiality] become especially salient to women seeking to improve their epistemic predicament’. Perhaps. But the vices reinforced by social norms might accordingly be *less* salient as foci for epistemic self-improvement. Astell was unusual—a visionary. Most contemporaries seeking self-improvement may have instead read novels, listened attentively at dinner, and aimed to absorb insights from men, who were considered epistemic superiors. (Indeed men actually were epistemic superiors in many domains because women were hamstrung by educational inequality.) Agitators aiming to improve the epistemic predicament of women may have campaigned for more tutoring or for permission to attend public lectures. These approximate gender-approved modes of self-betterment, such as absorbing information, sponge-like, from men, rather than gender-aberrant reforms, such as shedding submissiveness. Indeed, to many of Astell’s contemporaries, shedding submissiveness may have seemed degenerate, even to those seeking epistemic self-improvement.

Kidd writes, ‘women were, or would always become, marked by the “Feminine Vices”, like submissiveness and superficiality’. The term

'marked' has two connotations. The first is 'assigned', 'designated', or 'goes with'. In soccer and hockey, each attacker is marked by a separate defender, for example, driving lanes are 'marked for overtaking' and the third son is 'marked for the military'. (The first inherits the land; the second joins the clergy.) In this sense, 'women are "marked" as submissive' means 'women are normed as being submissive'. Secondly, 'marked' connotes that those traits stand out as conspicuous or salient.

I suggest women can be marked as submissive in the first sense, but not the second. Women's submissiveness can be non-salient even to those injured by that submissiveness or seeking to improve their situation. It can require acuity to clock epistemic vice and its social powers even though—or perhaps because—vice is so pervasive and injurious.

Acknowledgements

Many thanks to Jon Garthoff for helpful feedback on an earlier draft. This research was supported by an ACLS Fellowship from the American Council of Learned Societies.

Note

- 1 Kidd's claim is slightly stronger: Vice epistemology should proceed as a critical character epistemology; being attuned to epistemic vice rationally compels us towards the tenets of critical character epistemology. Cf. Mills's (2007) mapping the trajectory from naturalised, non-idealised epistemology to critical race epistemology.

References

- Astell, Mary (2002) *A Serious Proposal to the Ladies. Parts I and II* [1694]. Springborg (ed.), Broadview Literary Texts.
- Dillon, Robin (2012) "Critical Character Theory: Toward a Feminist Perspective on 'Vice' (and 'Virtue')". Crasnow and Superson (eds.), *Out from the Shadows*, Oxford University Press, pp. 83–114.
- Mills, Charles (2007) "White Ignorance". Sullivan and Tuana (eds.), *Race and Epistemologies of Ignorance*, State University of New York Press, pp. 11–38.

3d Ian James Kidd's Response to Commentaries

Rejoinder to Heather Battaly and Georgi Gardiner

IAN JAMES KIDD

I'm grateful to Heather and Georgi for their probing thoughts on my ongoing efforts to explore the possibilities for a critical character epistemology.

Blame and Structures

A critical character epistemologist sees epistemic agents as socially situated, their dispositions and activities being significantly shaped and often constituted by their social environments. I think those environments often tend to damage our epistemic character in various ways, this being the basic conviction motivating my concept of epistemic corruption. Heather asks how this relates to *blame*, an issue underplayed in my work so far. Certainly, the critical character epistemologist doesn't want to *rule out* our being responsible for the state of our epistemic characters. What they want, though, are complicated stories that issue in complex conditions: *certain* agents under *certain* conditions at or beyond *certain* points in their life and development can be judged responsible for the acquiring or retaining of at least some of their epistemic failings. To tell those stories, we need to get clearer on the sorts of explanations that are at work in accounts of epistemic corruption (causal or narrative ones, say?) A snag is that I think some of us are complicit in the deterioration of our own epistemic character. We can think of this as wilful, self-conscious epistemic self-corruption.

Heather also asks if social structures can themselves be vicious, in their sense of their bearing vices in their own right as well as facilitating their acquisition. I'm certainly happy to attribute epistemic vices to social structures and institutions and thereby expand the range of vice-bearers (Kidd 2021b). At the moment, though, I don't think social structures need to be vicious to be corrupting, partly because of what Margaret Gilbert called *divergence arguments* (see Gilbert 1989). I think individuals exist in complex dynamic relations with social structures and

institutions. Confronted with a dogmatic institution, one can acquiesce or resist in all sorts of ways shaped by individual character, situational pressures, interpersonal interactions, and so on. But settling this will require more thinking and some good empirical work.

Amelioration

In perhaps the hardest and most important question. Heather asks where those concerned about the amelioration of epistemic vices should start and on this I am divided. Certainly, one of the motivations for giving an account of epistemic corruption was to guide our critical thinking about the ways our epistemic characters get damaged: that is the ameliorative part of me. At the same time, I often fear that corrupting structures are too entrenched, or that any serious efforts to reform them may fail or backfire by supercharging our vices: this is the pessimistic and quietist part of me. In a sense, this is an uncertainty about the *nature* and *scale* of those ameliorative actions. Maybe we can reduce the incidence or frequency of severity of the vices of the mind—but that is an empirical issue about which I'm deeply ambivalent. Much will also depend on what is intended by *amelioration*. The modern tendency is to define this in terms of dramatic large-scale actions aimed at significant structural changes. But we should not rule out smaller-scale actions of a more modest character. After all, rapid and radical projects can also be sources or superchargers of epistemic vices. At the moment I am inclined to a pessimistic misanthropy: our personal and collective epistemic failings can be mitigated to certain limited degrees but never eradicated from what has come to be the human condition.

Salience

Georgi focuses on the different sorts and degrees of *salience* that epistemic vices can have, a rather neglected issue within the literature on character epistemology. It makes sense for vice epistemologists to offer analyses of specific vices without pressing onto questions about their salience for different individuals—up to a point. But at some point, we should start exploring the personal, situational, social, and cultural factors that shape the salience of different vices, where that includes diving into the historical work on vices (Kidd 2018, 2021a).

Georgi uses the term 'salient' in a tighter sense than I was, using it to mean attention-grabbing or cognitively prominent, whereas I used it in the looser sense of 'significant'. I agree that the narrower sense is more useful: a critical character epistemologist should say that the epistemic vices can have different sorts of *significance* for different people. More importantly, this can include two sorts of cases: first, pernicious forms of significance, like the misogynistic conviction that certain epistemic vices

ought to be especially significant to women since they are more prone to them—the claim being skewered by Astell in the remarks I quoted. Second, cases where certain epistemic vices *lack* the sorts of significance they ought to possess, even to the point of people being completely oblivious to them. Sometimes, this is because it suits certain powerful groups to conceal those vices from a collective understanding, a point Georgi makes using the case of chauvinist epistemic vices. In other cases, the obscuration of certain vices might be due mainly to historical contingencies in our inherited table of the vices—that being the main theme of my work in historical vice epistemology.

A key take-home from Georgi's remarks, and Heather's, is that our thinking about the nature, harms, origins, distribution, significance, and correctability of epistemic vices must be a multidisciplinary endeavour. Even at this early stage, vice epistemology has well-developed relations to virtue epistemology, social epistemology, feminist social philosophy, and areas of empirical psychology. Into the future, it should engage more with the social and historical and political dimensions of epistemic vice and with phenomena, like epistemic corruption, which force us to confront them. If it does, our ability to ameliorate them could match our ability to understand them.

References

- Gilbert, Margaret (1989) *On Social Facts* (Princeton, NJ: Princeton University Press).
- Kidd, Ian James (2018) 'Deep Epistemic Vices', *Journal of Philosophical Research* 43: 43–67.
- Kidd, Ian James (2021a) 'A Case for an Historical Vice Epistemology', *Humana. Mente* 14(39): 163–180.
- Kidd, Ian James (2021b) 'Epistemic Corruption and Political Institutions', Michael Hannon and Jeroen de Ridder (eds.), *The Routledge Handbook to Political Epistemology* (New York: Routledge), 347–358.

4 Narrowing the Scope of Virtue Epistemology

Neil Levy

There are many aims virtue epistemologists may seek to pursue. They may be interested in identifying and understanding dispositions or character traits that play important epistemic roles, and that is an aim that is surely legitimate. Situationist critiques of virtue theoretical approaches notwithstanding, it is very plausible that individuals differ in the degree to which they possess the kinds of dispositions widely taken as sufficient to classify them as having or lacking particular epistemic virtues. We may be interested in classifying people as virtuous or vicious, to some degree, and these classifications might aid our understanding. Working through careful discussions of open-mindedness or arrogance seems to have deepened my understanding of knowledge and belief. That's a significant payoff, which goes a long way to justify the enterprise.

However, one central aim of at least some virtue (and vice) epistemologists is meliorative.¹ That is, they are engaged in what Ballantyne (2019), and Roberts and Wood (2007) call *regulative epistemology*: epistemology that is designed to guide us in inquiry. I will argue that virtue epistemology is the wrong tool to employ in that enterprise, at least when regulative epistemology has the ambition to guide all or most of us in all or most of our intellectual lives. The virtues may have an important epistemic role to play, but only in circumscribed parts of our lives as enquiring beings. For the rest, we do better to focus on the epistemic environment. Moreover, it is largely by contributing to a knowledge-conducive epistemic environment that the virtues lead to better belief.

1

Virtue epistemology, in its regulative guise, aims to improve cognition by inculcating epistemic virtues. Correlatively, as a regulative enterprise vice epistemology counsels that we avoid the epistemic vices. Virtues and vices are character traits (and, perhaps, ways of thinking and attitudes too) which are, respectively, epistemically helpful and harmful. Either directly (by making us more or less responsive to evidence or criticism, say) or indirectly (by making us love truth or be indifferent to it, to read widely or to be incurious, say) they help or harm our functioning as

epistemic agents and lead us to have better or worse beliefs. Inculcating the virtues and helping us to avoid the vices is surely a worthy goal for a regulative epistemology.

At that level of abstraction, virtue epistemology sounds attractive. But when its proponents attempt to flesh out the details and show how it can be put into practice to improve our epistemic lives, they run into difficulties. In particular, when they neglect issues of scope – when they call on us to engage in responsible enquiry without regard to the topic or the specific expertise of the enquirer – they end up advocating strategies that cannot succeed.

Consider, for instance, how one virtue theoretical approach grapples with the so-called paradox of dogmatism. The paradox, in its original formulation due to Saul Kripke (2011), can be stated roughly as follows:

- 1 I know that p ; therefore p is true.
- 2 If p is true, then any apparent evidence e against p is misleading.
- 3 Since misleading evidence can be expected (all things considered) to make me worse off epistemically, I have good epistemic reason to ignore e .
- 4 Therefore, I should ignore any and all evidence against propositions I know.

The dogmatism paradox apparently licenses us to disregard evidence against any proposition we know to be true. While this is not, strictly speaking, paradoxical, it is uncomfortable insofar as it seems to warrant epistemically irresponsible behaviour. In virtue epistemological terms, it seems to warrant closed-mindedness, a paradigmatic epistemic vice.

Given these uncomfortable implications, the dogmatism paradox is usually seen as a puzzle to be solved. Solving it, in the sense meant here, would consist in identifying where it goes wrong, and thus why we shouldn't be closed-minded. But as Kripke himself notes, there are contexts in which dogmatism seems to be warranted:

[S]ometimes the dogmatic strategy is a rational one [...] Even when confronted with specific alleged evidence, I have sometimes ignored it although I did not know how to refute it. I once read part of a piece by a reasonably well-known person defending astrology [...] I was not in a position to refute specific claims but assumed that this piece was of no value.

(Kripke 2011, 49)

If Kripke's right, and dogmatism is not always epistemically irresponsible, we should not try to solve the paradox. Instead, we should seek criteria "to delineate cases when the dogmatic attitude is justified".

Of course, the demarcation of cases and cases is grist for a virtue theoretical mill. From its very inception in Aristotelian thought, virtue theorists have emphasized the need for good judgement to distinguish instances of courage from recklessness, or generosity from profligacy. But if Kripke is right that there are cases in which the dogmatic attitude is justified, the virtue epistemologist is in trouble. Dogmatism – or the dispositions that constitute dogmatic thinking – looks very much like the manifestation of the “archetypical epistemic vice”: closed-mindedness (Cassam 2018, 39). Indeed, Cassam concedes as much: “[i]f dogmatism isn’t an epistemic vice it is hard to see how closed-mindedness can be an epistemic vice” (109). In the abstract, the virtue theoretical response is obvious: distinguish the dispositions or traits, and maintain that holding firm in the face of arguments you can’t rebut, if it is ever appropriate, is not dogmatism but something else (just as charging into a fight you cannot win when there are more effective responses available is not courage but foolhardiness). This sort of response is open to the objection that it is merely verbal, insofar as it has about it more than a whiff of the suggestion that it turns not on differences between the dispositions or attitudes that are engaged but on the words, we use to describe these dispositions or attitudes. If we are to avoid the charge that the strategy is merely verbal, we need to distinguish cases in which dogmatism (or something very like it) is appropriate from those in which it is not, and – I will suggest – what distinguishes these cases is *not* the attitude but the context: we should be open-minded only in very restricted circumstances.

I will advance the case through a discussion of Cassam’s argument against dogmatism. A preliminary point: Cassam defines dogmatism narrowly. On his account, a person is dogmatic if (and only if) she ignores evidence that conflicts with her *doctrines*, where a “doctrine” is “a belief about the general character of the world, or some generally important aspect of the world, which bears the weight of many other beliefs” (106).² I will use “dogmatism” to refer to a policy of ignoring or refusing to consider what the believer themselves recognizes to be possible evidence against any (token) belief, whether the belief is central or peripheral to our epistemic network (or our network of cares). Thus, I can be dogmatic about whether the world is more than five minutes old or about whether I left my keys in my other trousers. I adopt this more expansive understanding of dogmatism to emphasize its scope: we routinely ignore certain kinds of apparent evidence against quite mundane propositions (and our doing so raises the issues at stake in the dogmatism paradox). Thus, a range of sceptical challenges to my belief that it is Wednesday, or that I slept at home last night, will be dismissed by me out of hand. I will take only certain sorts of (very rare) challenges to these beliefs at all seriously. My dogmatism about these mundane propositions is not different from my dogmatism about what Cassam

calls doctrines: there too I will take only certain sorts of – vanishingly rare – evidence against my beliefs seriously.

Cassam argues that the epistemic costs of dogmatism are higher than its advocates think and its benefits much smaller than they think. He also argues that the practice of exemplars of virtuous enquiry avoids dogmatism, contrary to what is often claimed. For evidence of its costs, he turns to twentieth-century history. According to Cassam, for example, Major-General Eli Zeira, the Director of Military Intelligence in Israel in 1973, was dogmatic in his belief that Egypt and Syria wouldn't attack, and his dogmatism led him to ignore the evidence against his belief. Even if Zeira's belief had turned out to be true, Cassam argues, his dogmatism prevented him from knowing that Egypt and Syria wouldn't attack because belief sustained by dogmatism rather than appropriate response to evidence is not justified. As Cassam puts it,

[w]here P is just a dogma to which S is attached in such a way that they would still be confident that P regardless of the evidence then S isn't guided by the evidence and doesn't have the right to be confident.

(110)³

Cassam concedes, nevertheless, that something in the ballpark of dogmatism is sometimes appropriate. Here he deploys the expected virtue theoretical strategy of distinguishing the traits, attitudes or dispositions involved in appropriate firmness from those that cause dogmatism. He develops this strategy with reference to Kuhn's contention that scientists are typically and appropriately dogmatic. Normal science, Kuhn argues, is science conducted within a scientific paradigm, where a "paradigm" is a set of taken-for-granted methodologies, findings, theories and exemplars of good scientific practice. Scientists are appropriately dogmatic inasmuch as they routinely reject scientific anomalies: findings or evidence that conflicts with the paradigm. Thus, for example, evidence of a genuine "saltation" in the evolutionary history of an organism will be regarded by biologists as spurious: evolution proceeds by small steps, not leaps, and there will be no change in phenotype that was not produced through a small change in its genotype or its environment. This looks like dogmatism, insofar as it involves the scientist regarding certain evidence as misleading simply on the grounds that it conflicts with what they take themselves to know.

Cassam denies it is dogmatism. A better label, he claims, is "firmness or tenacity" (113). It is, he argues, surely rational to respond to an anomaly by looking for ways to accommodate it or show that it is spurious.⁴ The scientist who abandoned her commitments too easily would not display open-mindedness, but rather intellectual "flaccidity". Indeed, it is built into the notion of having commitments that the person would not

revise them easily. Firmness is distinguished from dogmatism by the fact that the firm scientist will seek to defend her commitments in the face of objections, but is able and willing “to acknowledge fundamental flaws in established tools and beliefs, and abandon those tools and beliefs” (113). Firmness is the mean between flaccidity and dogmatism.

Having pointed to the epistemic costs of dogmatism and seen off, to his own satisfaction, the threat from the practice of scientists, Cassam praises the virtues of open-mindedness, even in the kinds of cases cited by Kripke. Kripke confesses he is unable to refute arguments he has encountered in favour of astrology and necromancy; he dogmatically ignores such arguments rather than attempt to refute them, and thereby protects his knowledge. Cassam denies that Kripke has any such inability: he can and should engage with these arguments. We should not fear misleading evidence, since it is (after all) misleading. It shows a vicious lack of self-trust to think that one will be taken in by such evidence. In fact, the lack of confidence that dogmatism in the face of misleading evidence manifests is *itself* a threat to knowledge, since (in Cassam’s view) knowledge requires confidence.

Instead, one should be confident in one’s ability to confront misleading evidence. One can and should figure out where arguments in favour of astrology go wrong, or – when technical expertise one lacks is required – consult the experts, and work out which of disagreeing experts is more likely to be right. The more dubious the theory, the easier it is to dismiss, so there’s no real danger that we might lose knowledge of claims like “necromancy is bunk”. Conspiracy theories call, Cassam says, for a “serious response”: a rebuttal, not a mere denial. We can give such a response: most people are perfectly capable of checking and assessing what they hear and read. Consider Holocaust denial, which serves as Cassam’s central example in this chapter. If we encounter the poisonous claims of a David Irving, we should do our due diligence. If we Google him, we will discover that he was found by a court to have deliberately misrepresented historical evidence to promote Holocaust denial and that his interpretation of key historical documents has been discredited by credible historians.

The appropriately firm agent exhibits intellectual firmness in the face of David Irving-style conspiracy or the superstitions discussed by Kripke. Rather than abandon her beliefs (that the Holocaust happened; that necromancy is bunk; and so on) she confronts arguments and evidence that are purported to refute them, confident they are spurious. The flaccid person would abandon the belief in the face of arguments against it. The dogmatic person would dismiss the evidence out of hand. The firm person confronts it and disarms it.

I don’t think any of this is remotely satisfactory. Cassam succeeds neither in making a convincing case for the claim that we may secure our knowledge by confronting misleading evidence, and he mischaracterizes

the attitudes of the scientist. Both the scientist and the layperson do and should adopt a stance that counts as dogmatic by Cassam's lights: ignoring contrary evidence and refusing to budge even when she cannot accommodate evidence she recognizes to be anomalous. If the scientist, who devotes her professional life to the careful examination of evidence, must be dogmatic with regard to much of it, then the ordinary person (who has much less time, and far fewer tools, for such examination) is well advised to take the same approach. As Kripke suggests, the real trick is not deciding *whether* to be dogmatic, but instead identifying *when* dogmatism is the appropriate response.

Cassam's arguments against dogmatism have three main elements: the costs of dogmatism, the behaviour of the scientist, and the benefits of confronting misleading evidence. I won't address his arguments one by one, under the same headings, because the links between these topics are too intimate. I will aim to show that the *all things considered* costs of dogmatism are very much lower than Cassam claims. Indeed, when it is appropriately deployed, the all things considered costs of dogmatism are negative – that is, it has more benefits than costs. Showing that that's the case depends on showing that the benefits of confronting misleading evidence are very much lower than Cassam claims. I will proceed by directly assessing the prospects of doing what Cassam calls on us to do: rebutting misleading evidence. I will also argue that Cassam mischaracterizes the behaviour of scientists, who are (and should be) far more dogmatic than he realizes.

Let's begin with an assessment of the epistemic costs and benefits of dogmatism. Take climate change, for example. Every single day, someone claims to have evidence that is incompatible with the basic outlines of the consensus position on anthropogenic global warming. Almost as frequently, the claim is made by someone with apparent scientific expertise and data to back it up. Just today (as I write these words), I came across a book called *The Rise and Fall of the Carbon Dioxide Theory of Climate Change*, which apparently defends the well-known "sceptical" hypothesis that climate change is caused by fluctuations in solar energy. The book is published by Springer, a reputable publisher. The author, Rex Fleming, has a PhD in atmospheric science from the University of Michigan and is an elected fellow of the American Association for the Advancement of Science. His publications, on a variety of scientific topics, include recent papers in the *Journal of the Atmospheric Sciences* (impact factor 3.159) and *Environmental Earth Sciences* (impact factor 1.871).

According to Fleming's own website, the unique insight his book identifies is

the failure of the Schwarzschild radiation integrations to maintain the CO₂ longwave radiation intensity achieved in the surface warming by H₂O and CO₂. The resultant Planck radiation intensity is

severely depleted in the upper atmosphere. The result is the CO₂ molecules merely pass their remaining small residual heat to space un-impeded.

If climate science is representative of those topics on which the epistemically virtuous response to misleading evidence is rebuttal (rather than one of those on which the virtuous response is to identify the best expert and defer to them), having read this brief description puts you under an epistemic obligation: if you are to retain the knowledge that climate change is very largely or exclusively caused by human activity, you must rebut the claim that “CO₂ molecules merely pass their remaining small residual heat to space un-impeded”. Is this really something you can do? I am sceptical that anyone who reads this chapter will have the requisite expertise to responsibly assess this claim. I have little idea what the words above mean, beyond the fact that they are taken to entail that CO₂ molecules have heat and that heat dissipates without effects on the climate. I could, of course, google “Schwarzschild radiation”, “integrations”, “longwave radiation intensity” (or should that be “longwave” and “radiation intensity”?) and try to discover what the phrases mean, preliminary to assessing them, but I strongly suspect that it would take me not hours but days to get a glimmering of understanding of what these phrases mean. Worse, doing so will *worsen* my epistemic position, not improve it: I will then have a better understanding of Fleming’s claims against the AGW consensus, not a way of rebutting these claims.

While I think (perhaps optimistically) that if I devoted some days to the project, I could come to a reasonably clear understanding of Fleming’s main claims, I doubt I would *ever* be in a position to rebut them, no matter how hard I worked at it. Frankly, I lack the maths, and without the maths, it’s usually impossible to get a sufficiently deep grasp of the sciences to be able to assess the claims made in the technical literature. Perhaps I could acquire the maths? Perhaps, but even if I already had it, coming to be in a position to rebut Fleming is *already* a project requiring literally thousands of hours of immersion in the technical literature. If I can come to be in a position responsibly to rebut Fleming’s claims (the claims, recall, of someone with a PhD and publications in climate science, as well as in the development of predictive mathematical models) it will be acquiring a good chunk of the expertise of the climate scientist. How much expertise would I need? This is not a question I can answer, because it would take possession of the very expertise I lack to assess it. Confronting misleading arguments like this one will require some degree of genuine expertise, depending on how complex the arguments are and how subtle the errors involved might be. Sometimes, possession of a good four-year degree in the subject will be enough. Sometimes PhD-level expertise, or perhaps even better, will be required to identify

errors and rebut the claims.⁵ As we will see, sometimes genuine high-level expertise isn't sufficient to rebut misleading evidence.

Once we recognize just how high are the barriers to rebuttal, for an ordinary person (even one like me, with access to university libraries and extensive research experience), it is immediately apparent that few of us can ever rebut sophisticated climate change denial (at very most, a very few of us will have time enough to gain expertise only in a very circumscribed area of specialist knowledge). I suspect Cassam would agree, given that he notes knowledge of physics or engineering might be necessary to refute the 9/11 conspiracy theorist, but that it would be "unreasonable" (117) to expect ordinary people to acquire such knowledge. Instead, he suggests, we should refute the misleading evidence "by consulting experts and working out who is most likely to be right". Before discussing this alternative method of rebutting the climate sceptic or the Holocaust denier, let me turn to the expert herself. Surely, *she* can reasonably be expected to rebut the sceptic?

Of course, some scientists may be in a position very rapidly to see where Fleming has gone wrong. They may have sufficient expertise in Schwarzschild radiation integrations, and so forth, to assess and dismiss Fleming's claims by reading perhaps no more than a part of his book, or even the summary I have quoted above. But it bears emphasising that often the number of scientists who can identify the problems rapidly will be quite low. Science is highly specialized, and it is frequently the case that scientists lack the specialized expertise to assess claims made in their general, but not specific, area. For instance (and here I cite a real example from my own experience) a neuroscientist may be quite at a loss when it comes to claims about the functional role of a particular brain region, even though they specialize in that very brain region, because their interest is in gene expression and the development of that region, and not in what it does.

Many neuroscientists who lack the specific expertise required to assess a claim within their general area can come to acquire it relatively rapidly. How rapidly will differ from case to case: a neuroscientist who has specialized in gene expression may not be in a better position to understand the functional role of a brain region than a mathematician, say: her path to specialization may not even have involved many undergraduate courses in common with the cognitive neuroscientist. In some cases, only a few days might be required to acquire sufficient expertise to identify and dismiss the cranks. Even for those who have *specific* expertise, some investment of time is required to rebut misleading claims: in the best of cases, the time taken to read at least a little of (for example) Fleming's arguments. Given that there are many spurious claims made, this is an expenditure of time and effort most will avoid paying. Scientists are keen to get on with their own research. They want to read useful material, material that advances their work (often by challenging it)

not waste their time on identifying confusions. While some will happily spend their downtime in reading and refuting kooks and cranks, many will regard themselves as having more important things to do. Given the investment of time required and the limits of specialized expertise, even if Cassam is correct that they have an obligation to rebut wild claims, they will be able to address only a tiny proportion of these claims. There simply isn't enough time for even the most dedicated conspiracy rebutter to do more.

It might be objected that neuroscience and atmospheric science, with their demands for technical expertise and their heavy reliance on advanced mathematics, are unusual in requiring a very heavy investment of time for responsible rebuttal, or at any rate that there are other areas in which sufficient expertise can be acquired quite rapidly, from a standing start. As already mentioned, Cassam's prime example in Chapter 5 of his book is Holocaust denial, and specifically the claims of David Irving. If we want to assess Irving's claims for ourselves, Cassam maintains, it is sufficient to read Richard J. Evans' *Telling Lies about Hitler*. There we will see Irving's lies "brilliantly exposed" (114). Perhaps history is unlike science: whereas in the former possession of demanding field-specific technical expertise is required to adjudicate debates, even when one side is mendacious, in history we need only common sense and a good book to see through the lies.

I think this claim very seriously underestimates the degree to which historians possess – and must deploy – field-specific expertise. In fact, just as in science, the expertise possessed by a historian is not merely specific to the field of history, but specific to a historical period and perhaps much more specific than that. To expose Irving, it was necessary to possess a wide range of background knowledge – concerning how the German state worked, about the jargon of bureaucrats, about the role of different members of Hitler's inner circle, and so on – not merely apply common sense. Evans' background knowledge, as well as the specific interpretive tools of the historian, cannot themselves be conveyed to the non-expert reader. Instead, the reader can only be given the rough outline of his reasons for certainty that Irving is distorting the historical record.

Consider, for illustration, Naomi Wolf's recent public embarrassment over her new book (Wolf 2019). In *Outrage*, Wolf argues that persecution and prosecution of "sodomy" increased significantly after 1857. A key piece of evidence for her claim was the appearance in court records of the phrase "death recorded". Wolf interpreted the phrase as meaning that the person had been sentenced to death. In actual fact, it was used for a nominal death sentence: one which would not be carried out. Wolf had, of course, done a great deal of research for her book. In fact, it was based on her Oxford University doctoral dissertation, supervised by an expert in nineteenth-century English literature. Wolf came in for

a great deal of derision for her supposed failure to fact-check her work. But Wolf had good reason to be confident in her work: not only had it passed through the Oxford examination, but she had enlisted the aid of Dame Helena Kennedy, a prominent human rights lawyer, to check her interpretation of the law. Kennedy interpreted “death recorded” in the same way Wolf had (Kennedy 2019).

This episode, which is by no means an isolated one,⁶ demonstrates how much-specialized knowledge is required to interpret historical documents. One needs not the expertise of a historian, but the *specialized* expertise of the historian who works on that period specifically, and on particular aspects of that period at that (Wulf 2019). Surely it takes less expertise to assess competing accounts between duelling experts than it does to generate these accounts – a historian of, say, modern Europe (but who lacks the expertise specific to the Nazi period) is better able to adjudicate between Irving and Evans purely on the basis of the arguments and evidence each presents than am I – but such adjudication will still require some degree of genuine expertise (just as we need a degree of genuine expertise to assess the claims of those who put their skills in the service of climate denial). Adjudicating on a debate between two people who possess genuine expertise is difficult, and this remains true even if one of them is mendacious (Irving possesses genuine expertise: prior to becoming a full-blown Holocaust denier, Irving published several books, one of which is still well regarded. This expertise gives him the tools to distort history in a way that it takes genuine expertise to expose). Evans may indeed brilliantly expose Irving’s lies, but it’s not because of our capacity to assess the dispute between Evans and Irving that we accept the former’s account.

If I’m right that Cassam seriously overestimates the capacity of the ordinary intelligent person to adjudicate the Evans/Irving dispute, then his less-demanding prescription for those of us who lack “the time, energy, or intellectual resources” to read books like Evans’ must fail abjectly, at least if it is understood in the way he understands it. Cassam advises us to turn to Google and Wikipedia. There we will learn (for instance) that Irving was found to have deliberately distorted the historical evidence by a British court, and that the interpretation of his evidence has been discredited. Of course, Cassam is quite right that those who search will find these claims reported, but why should they accept either that Wikipedia accurately reports the court’s judgement or – more particularly – that the court was correct? After all, some more googling will lead to websites claiming that the court got it wrong. It cannot be the case that we ought to accept the claims we read on Wikipedia because it can do what academic historians cannot: convey to readers not only the findings of historians but also the entire intellectual edifice that justifies these findings. Again, if we attempt to settle the issue for ourselves *on the basis of the evidence and arguments presented*, Wikipedia and Google will let us

down (in a moment we will see that reading about the court's judgement *does* provide us with grounds for siding with Evans and not Irving, but not because of the arguments Wikipedia presents).

Notoriously, those who turn to google to conduct research, in the manner Cassam recommends, often end up with more distorted views. After all, if one is carrying out one's research conscientiously (in a way that manifests the intellectual virtues like open-mindedness) one had better give a fair hearing to both sides. Doing that, though, soon leads to a thicket of claims and counter-claims, few of which the non-expert consumer is in a position to assess for herself. Do vaccines cause autism? Well, a peer-reviewed paper published in the prestigious journal *The Lancet* made that claim. That paper was later retracted and found to be fraudulent. But isn't that exactly what you would expect from a journal system that relies heavily on industry funding and is therefore reluctant to criticize it? If you think that that's paranoid, recall how Elsevier – the publisher of *The Lancet* – produced six fake journals to deceptively present industry-friendly content as though it appeared in peer-reviewed articles (Hutson 2009). At *best*, the non-expert who attempts to give both sides a fair hearing ends up aware of a range of conflicting claims (e.g., about the efficacy and safety of vaccines; about the behaviour of people on each side of the debate; about the role of drug companies, and so on) which she is no position to assess for herself, and therefore comes to be in a worse epistemic position (Levy 2006). Even if she comes or continues to believe the truth (that vaccines are safe and effective, say) she may nevertheless lack knowledge. Alternatively, in the face of her inability to assess the competing claims, she may become agnostic, thereby losing knowledge and belief.

Naomi Wolf's experience provides several unhappy examples of what Ballantyne (2019) calls "epistemic trespassing", where someone with genuine expertise in one field takes themselves to have sufficient expertise to engage seriously with another. Wolf and her supervisor took their expertise in British nineteenth-century literature to equip them to interpret nineteenth-century British legal texts; Dame Kennedy took her expertise in contemporary law to equip her to interpret the law of the past. Epistemic trespassing can have unhappy consequences even when, as seems to be true in this case, the trespassers have sufficiently closely related expertise to be unaware of themselves as trespassing. Ballantyne gives examples of successful transfer of skills from one domain to another, but given the low success rate, the epistemically responsible agent will refrain from such transfer, at least unaided by someone who is genuinely at home in the field (note that in Ballantyne's principal case of successful transfer of skills across domains – in which high-school students attempted to explain historical events on the basis of fragmentary evidence – those who lacked skill in the target domain did surprisingly well, but were nevertheless and entirely unsurprisingly very

significantly outperformed by those who possessed such skills). Abandon dogmatism and face the risks of epistemic trespassing, and losing knowledge. But it is not just epistemic trespassers who may responsibly be dogmatic. Scientists, working within the domain of their own expertise, sometimes confront findings that they cannot explain. They are often, rightly, dogmatic in the face of such findings: simply setting them aside as anomalies.

In other words, Cassam mischaracterizes how scientists behave in the face of anomalies. According to him, they exhibit firmness: neither folding in the face of anomalies nor dogmatically ignoring them. Responsible scientists, he claims, are ready “to acknowledge fundamental flaws in established tools and beliefs, and abandon those tools and beliefs” (113). But that does not accurately describe the scientific practice. When scientists are in possession of a research paradigm that unifies a great deal of disparate work and has proven to have predictive power, they do not abandon it or even acknowledge flaws (let alone fundamental flaws) in the face of anomalies. They may not even pause to examine anomalies. When scientists encounter anomalies they can’t explain, they often set them aside, in the expectation that the future advance of science will accommodate the finding.

Science is in fact littered with examples of scientists holding fast in this kind of way. Consider how Darwin and those who followed him reacted to Kelvin’s careful work on the age of the Earth. The estimated range he produced was, as Darwin recognized, far too short for the diversity of life to be explained by natural selection. Despite recognizing that he was unable to refute Kelvin’s findings, Darwin refused to abandon his theory. Of course, new evidence entirely vindicated Darwin, but he held fast long before he was able to cite this evidence himself (Lewis 2002). Scientists are much more dogmatic than Cassam suggests, and this is epistemically appropriate for reasons Kuhn gave: because abandoning a paradigm prematurely leaves us unable even to recognize anomalies, let alone explain them, and because entrenched paradigms usually prove able to explain the apparently anomalous in the end (Kuhn, 1970).

As we saw, Cassam argues that the costs of dogmatism can be high. Indeed, they can: just as there are plentiful examples of scientists who held fast in the face of anomalies, subsequently to be vindicated, there are plentiful examples of dogmatism in the face of anomalies that proved intractable and ultimately could only be explained by a new paradigm. Consider the medical community’s dogmatism in the face of evidence that antibiotics were a successful treatment for stomach ulcers. This anomaly was ultimately explained only when the stomach acid theory of ulcer formation was rejected, in favour of a bacterial hypothesis; in the meantime, doctors were sufficiently convinced of their false theory to support fines for doctors who used antibiotics as a treatment (Zollman 2010). If I am correct, however, we should be dogmatic in the face of

anomalies, because we can usually expect the anomaly to be eventually explicable within the existing paradigm, and we do worse to reject the paradigm without a viable alternative available in any case, then the costs should be paid. The benefits of dogmatism are routinely higher than the costs.

If all this is correct, then Kripke was right: the question is not *whether* to be dogmatic, but *when*. In what contexts should we dismiss evidence and arguments offered against our prior beliefs? The answer, I claim, is very often. When we have acquired our beliefs through testimony, and that testimony is sufficiently good that our belief is a good candidate for knowledge, we ought to stand fast in the face of anomaly or evidence against our belief, unless the source of the evidence has the same kind of standing as the original source of testimony.

Laypeople (and of course, we are all laypeople with regard to most areas of knowledge) acquire their beliefs about specialist topics by testimony, explicit or implicit (beliefs are acquired by implicit testimony when they are based on claims that are not asserted but presupposed or implicated; see Levy (2019) for discussion). These beliefs are good candidates for knowledge when the (ultimate) source of the testimony is, or is representative of, the appropriately constituted epistemic authorities. I cannot make even the beginnings of a proper start on an account of what properties an epistemic authority has and what makes such an authority properly constituted. I want to highlight just one – central and very important – authority-conferring property: the social constitution of knowledge.

On most specialized topics, at least, an epistemic authority is not, and does not speak as or for, an individual. Rather, the authority is or speaks for a group, and that's no accident: knowledge of specialized subjects is the product of a deep division of cognitive labour. Science is of course the paradigm of such a division of labour, partially conflictual and deeply cooperative. Cooperation (mostly) characterizes relations at the level of the lab, which is to a large extent the unit of scientific production; relations between labs are characterized by both conflict and cooperation. Labs seek to refute one another, but they take one another's data and results (largely) on trust. Conflict and cooperation are institutionalized in peer review and (increasingly) in post-publication review of results. In disciplines beyond the sciences, the unit of production is often individual, but knowledge arises through conflict and cooperation across individuals just as much as in the sciences. Perhaps there are exceptions: perhaps there are important areas of knowledge that do not arise from conflict and cooperation across individuals and groups. Certainly, the degree to which knowledge is social differs from topic to topic. For most of what we know, however, beyond the deliverances of our senses (perhaps) such social relations are very important, and for all or almost all of the topics with which regulative epistemology

is concerned (knowledge of science, of history, of current affairs, of the state of the economy, of policies, of facts about public figures, and so on) it is very deeply social.

The proper epistemic authorities are hooked into these social networks in the right way, such that they can report the consensus view (when there is one) on a topic. The representative scientific bodies are hooked into these networks, and report a consensus of their members via press releases and talking to journalists. It is (in part) because bodies such as the IPCC, the American Association for the Advancement of Science, The American Geophysical Union and a host of other such bodies endorse the consensus on climate change that we know it to be true. It is because the AMA endorses vaccines that we know them to be safe and effective. In the face of challenges to beliefs like this, we are rightly dogmatic. When we have acquired beliefs via testimony from the epistemic authorities, we rightly dismiss arguments or evidence against these beliefs, unless these arguments/evidence come from the same authorities (and are presented by them as representing a challenge to our beliefs). We can thus retain knowledge. If we are not dogmatic, we run a very large risk of losing it, and are vanishingly unlikely to improve our epistemic position with regard to the beliefs in question.

As we have seen, scientists themselves – those who help constitute epistemic authorities – often are rightly dogmatic in the face of anomaly. Most scientists most of the time are in the same position as the layperson with regard to such challenges, so that's not a surprising result. The doctor who is presented with evidence that vaccines cause autism may be entirely unable to rebut the argument, but she may rightly shrug her shoulders. She should defer, just as we should, unless this is her precise speciality. Even if it is her precise speciality, rebutting the evidence may be a waste of her time. She may use heuristics to parse whether the argument is worth granting even *prima facie* plausibility, ignoring the evidence of the clearly unqualified. As for the rest, when she is presented with a *prima facie* credible argument by someone who is in a position to knowledgeably advance such an argument, she *still* need not engage. At most, there is an obligation for *someone* to take the argument seriously. The rightful scope of dogmatism is very broad.

The conduct of enquiry does, for all that, require something like open-minded enquiry. But the scope of such enquiry is very narrow. The conscientious scientist takes challenges (from those with the right credentials or who pass other stringent tests for expertise) to the hypotheses she is developing in her precise area of expertise seriously. She is not dogmatic with regard to them. Perhaps virtue epistemology well describes the dispositions and attitudes she displays in this very circumscribed area (I take no stand on that question). For the most part, however, she should be dogmatic. So should the layperson be dogmatic on those questions on which she lacks expertise.

Virtue epistemology as regulative epistemology therefore may have a target, but it is a narrow one and – correspondingly – its pretensions to guide us in our epistemic lives should be considerably deflated. Of course, I have focused here on dogmatism and the corresponding virtue of open-mindedness. But the point generalizes, I believe. The epistemic virtues are dispositions, character traits or attitudes that enable us to think for ourselves. And that’s exactly what we shouldn’t be doing. We should be deferring (manifesting, if anything, excessive gullibility by the standards of virtue epistemology).

Let me finish with an important caveat. There is one area of our lives that is very important and in which appropriate epistemic agency may depend on the virtues: our personal lives. Our friendships, our intimate relations, our relations to our co-workers are also areas in which we exercise epistemic agency. In our personal lives, too, knowledge depends *very* heavily on testimony, but properly constituted epistemic authorities are much rarer, and we may be called on to adjudicate between conflicting sources of testimony or to weigh instances of it for plausibility. Here the scope for specialized knowledge is much narrower (though it is plausible that even here we should give scientific claims much greater weight than we do, rather than rely on folk psychology).⁷ Perhaps virtue epistemology as regulative project has important work to do in this region. But its ambitions to improve public discourse or to bring us to have better beliefs about matters of public importance are probably misplaced.

2 Conclusion

Regulative epistemology is, arguably, the most important branch of epistemology. It matters what people believe, and the project of making us more responsive to good evidence is an important one. It is unlikely, however, that virtue epistemology has a large role to play in regulative epistemology. The more important the belief – the more it is a belief that is relevant to our functioning in the public sphere – the less it matters whether we display the epistemic virtues. It is only in the narrow sphere of our own specialist expertise and our private lives that we ought to display the virtues. And this is the case because it is only in these spheres that we ought to be thinking for ourselves. For the rest, we ought to be deferring.

But isn’t appropriate deference itself the manifestation of a virtue? Perhaps: perhaps there is a virtue of (extreme) epistemic humility that such deference displays. Epistemic humility might be a kind of master virtue; the virtue that underlies our appropriate activity as epistemic agents. The available evidence concerning when and why agents defer to properly constituted authority sits uneasily with this suggestion, however. Rather, the available evidence suggests that the psychological processes underlying deference to bad authority are identical to the processes underlying deference

to good (Levy, Forthcoming). In both cases, deference is responsive to cues the person (implicitly) takes to be evidence of trustworthiness, and in both cases, this responsiveness is rationally appropriate. If we are to make people better responsive to *reliable* authority, we don't need to change people; not at the level of their epistemic dispositions, at any rate. Rather, we must ensure that the cues to which they respond are appropriately matched to the filters they deploy, and that's a matter of changing society. We must ensure that science is bipartisan, for instance, so that reliable measures pass the tests everyone uses; we must ensure that the epistemic environment is unpolluted, and so on. There is a lot of work for regulative epistemology to do, but this is work on society not on the individual.

Notes

- 1 For the purposes of this paper, at very least, I won't distinguish between virtue and vice epistemology. It's not quite true that they mirror one another, in that the virtues identified by the first are just the absence of the vices identified by the second, and vice-versa (such that we can restate the conclusions of each in the vocabulary of the other), nor is it quite true that the tools each uses are more or less identical to the tools of the other, but it is near enough to true for me to set the remainder aside. It should be clear that I have in mind responsibilist virtue and vice theory, of course: virtue reliabilists need not be interested in character traits at all and virtue reliabilism is better suited to explaining the simple cases of knowledge arising from faculties functioning as designed in the environments for which they are appropriate than the complex cases which cause dissent and which motivate the regulative project in the first place.
- 2 Here Cassam is quoting Roberts and Wood (2007, 194). Their notion of a doctrine is uncomfortably close to the idea of what is sometimes called a *hinge proposition*; uncomfortably close, because hinge propositions are often taken to be immune to doubt.
- 3 We shall see later that this kind of sensitivity principle should be rejected by a deeply social response to the dogmatism paradox.
- 4 In making this point, Cassam once again relies on Roberts and Wood (2007, 183–185).
- 5 To bring home just how difficult it is for those who lack genuine and deep expertise to assess controversial scientific claims for themselves, let me use the example of implicit bias, discussion of which occupies the bulk of Chapter 7 of *Vices of the Mind*. In that chapter, Cassam makes a number of (suitably) qualified claims about implicit attitudes and the implicit association test. For example, summing up some pages of discussion he concludes “there is no empirical justification for the view that it is impossible to improve one's implicit attitudes, or there is nothing that a person can do to change. Self-improvement in this area is possible and there are specific means by which it is possible, given the requisite levels of awareness, motivation, and skill” (173). I've had implicit attitudes as a central research interest for more than a decade (see (Levy 2017, 2016, 2015, 2014a, 2014b)). Yet I'm confident neither that Cassam's cautiously phrased claims are true nor that they are false.
- 6 Wulf (2019) gives the example of Cokie Roberts' claim, on NPR, that contemporary historians writing about abortion in the nineteenth century were distorting the historical record, on the basis that contrary to their claims

there were no advertisements for abortion services in nineteenth century newspapers. In fact, Roberts simply lacked the expertise to identify the relevant advertisements, which were plentiful.

- 7 For example, most people believe that memories are a highly reliable snapshot of events. But there is extensive evidence that memories are reconstructed, rather than simply recalled (such that features of the context of recall may affect the content of what is recalled) and that even important events may be remembered inaccurately. In particular, memory is easily contaminated: extraneous or false information may be advertently or inadvertently introduced, with the result that the person confuses information introduced later with information available only earlier. This kind of contamination explains some instances of misidentification of suspects by witnesses: they mistake the person in the police line-up or the mug shot with the person who committed the crime, for instance (Wixted et al. 2018). The folk belief that memory is reliable probably makes such contamination more likely.

References

- Ballantyne, N., 2019. *Knowing Our Limits*. Oxford University Press, Oxford.
- Cassam, Q., 2018. *Vices of the Mind: From the Intellectual to the Political*. Oxford University Press.
- Hutson, S., 2009. Publication of Fake Journals Raises Ethical Questions. *Nature Medicine* 15, 598–598. <https://doi.org/10.1038/nm0609-598a>
- Kennedy, H., 2019. I Fact-Checked Naomi Wolf's New Book on Gay Rights – The Mockery She Has Faced for One Error Disguises the Real Outrage We Should Feel. *The Independent*.
- Kripke, S.A., 2011. *Philosophical Troubles: Collected Papers*. Oxford University Press.
- Kuhn, T.S., 1970. *The Structure of Scientific Revolutions*. University of Chicago Press.
- Levy, N. Forthcoming. What does the CRT measure? Poor Performance May Arise from Rational Processes. *Philosophical Psychology*.
- Levy, N., 2019. Nudge, Nudge, Wink, Wink: Nudging Is Giving Reasons. *Ergo: An Open Access Journal of Philosophy* 6. <https://doi.org/10.3998/ergo.12405314.0006.010>
- Levy, N., 2017. Am I a Racist? Implicit Bias and the Ascription of Racism. *Philosophical Quarterly* 67, 534–551. <https://doi.org/10.1093/pq/pqw070>
- Levy, N., 2016. Implicit Bias and Moral Responsibility: Probing the Data. *Philosophy and Phenomenological Research* 93, 3–26. <https://doi.org/10.1111/phpr.12352>
- Levy, N., 2015. Neither Fish Nor Fowl: Implicit Attitudes as Patchy Endorsements. *Noûs* 49, 800–823. <https://doi.org/10.1111/nous.12074>
- Levy, N., 2014a. *Consciousness and Moral Responsibility*. Oxford University Press.
- Levy, N., 2014b. Consciousness, Implicit Attitudes and Moral Responsibility. *Noûs* 48, 21–40. <https://doi.org/10.1111/j.1468-0068.2011.00853.x>
- Levy, N., 2006. Open-Mindedness and the Duty to Gather Evidence. *Public Affairs Quarterly* 20, 55–66.
- Lewis, C., 2002. *The Dating Game: One Man's Search for the Age of the Earth*. Cambridge University Press.

- Roberts, R.C., Wood, W.J., 2007. *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford University Press.
- Wixted, J.T., Mickes, L., Fisher, R.P., 2018. Rethinking the Reliability of Eyewitness Memory. *Perspectives on Psychological Science* 13, 324–335. <https://doi.org/10.1177/1745691617734878>
- Wolf, N., 2019. *Outrages: Sex, Censorship and the Criminalisation of Love*. Virago Press Ltd.
- Wulf, K., 2019. Perspective | What Naomi Wolf and Cokie Roberts Teach Us about the Need for Historians. *Washington Post*.
- Zollman, K., 2010. The Epistemic Benefit of Transient Diversity. *Erkenntnis* 72, 17–35.

T&F Proofs – Not for Distribution

4b Commentary from Steven Bland

Expanding Our Notion of Virtue: A Commentary on Neil Levy's "Narrowing the Scope of Virtue Epistemology"

STEVEN BLAND

I am deeply sympathetic with Neil Levy's vision of a thoroughly social epistemology. With Levy, I believe that responsibilist virtues, such as open-mindedness and intellectual autonomy, are not as robustly beneficial as virtue epistemologists believe. Indeed, given that these dispositions systematically interfere with the generation and transmission of knowledge when manifested by individuals in collectivist contexts, it seems that they can be thoroughly deleterious. This is even more likely if we accept, as I think we should, what Levy calls the "social constitution of knowledge".

In Levy's view, the contextual instability of responsibilist virtues (and vices) presents two problems for the regulative ambitions of epistemic virtue theories. The first I will call the *problem of scope insensitivity*: virtue theories are insufficiently sensitive to the scope of their norms. If epistemic dispositions are beneficial in some environments and detrimental in others, then virtue theoretic norms ought to make this explicit; otherwise, their guidance will have mixed results, at best. For example, Levy argues that the unqualified manifestation of open-mindedness leads to knowledge loss, resource misallocation, and epistemic trespassing. For this reason, he claims that we should be dogmatically closed-minded with respect to any question on which we lack the requisite expertise. And he points out that most of the questions we care about belong in this category. This leads to the *problem of scope overreach*: given the narrow scope of most epistemic virtues, our regulative goals are better achieved by reforming the environments in which we think than the ways in which we think. While I am fully on board with Levy's first objection, I think the second objection might itself be somewhat of an overreach.

My hesitation is not with the situationist insight that our cognitive lives can be improved by reforming the contexts in which we live

them; indeed, I think this point is both true and important. Rather, it is with the framing of this position as an *alternative* to agent-centred approaches. This framing feeds into the person-situation debate that has occupied psychology and philosophy over the last few decades. This debate has given rise not only to divergent descriptive programs, but to alternative ameliorative approaches: one focussing on cultivating better dispositions, and the other on designing better environments. I call these, following Trout (2005), *inside* and *outside* strategies, respectively.

It seems to me that this debate is largely over, not because one side has proven itself superior to the other, but because it is based on a false dichotomy. Most psychologists now agree that personal and situational factors are not independent causal vectors, but entwined forces whose *interactions* are the principal source of human behaviour (Kihlstrom 2013). Furthermore, our environments and dispositions are *reciprocally determined*: each exerts a strong influence over the other. The appropriate response to this state of affairs is not to narrow the scope of virtue epistemology, but to broaden our conception of epistemic virtues (and vices). On one hand, we should understand virtues as being *situationally embedded*, that is, as being systematically dependent on environmental factors for their epistemic status, manifestation and cultivation (Skorburg & Alfano 2019). On the other, we should recognize that there are virtues whose value consists in their tendency to promote benign cognitive environments. We might call these *embedding virtues*. This broadening of our notion of epistemic virtues uncovers regulative strategies that defy straightforward classification as being either inside or outside. *Outside-in* strategies scaffold environments to promote virtuous habits; *inside-out* strategies cultivate habits of scaffolding benign environments (Bland, this volume). The upshot of this interactionist view is that cognitive dispositions and environments must be *coordinated*, rather than improved in isolation.

So, while I agree with Levy that the social transmission of knowledge often requires dogmatic deference, and that this process can be improved by making changes to the environments in which knowledge is socially transmitted, I would not want to overlook the role that personal dispositions can play in this project. Our deference is not capricious, but naturally guided by a set of heuristics: we preferentially defer to successful and prestigious individuals, and we often defer to the majority. While our reliance on these heuristics leads to cognitive distortions – success bias; prestige bias; conformity bias – it’s been essential to the cognitive success of our social species (Henrich 2016). It serves us less well, however, in our increasingly digital environments, where our perceptions of success and prestige get distorted, and the pressure to conform can be overwhelming, leaving us susceptible to manipulation by online “influencers”. We could seek to remedy this situation by cultivating greater epistemic discernment, but Levy apparently favours the outside strategy

of designing digital environments that better fit our evolved heuristics. I'm inclined to think that this would be a more successful approach, *if* it were to be widely adopted. I despair of the prospects of implementing this plan, however. Instead, we might teach individuals to curate their own digital environments so that they needn't be hyper-discerning when presented with information online. In short, epistemic virtues need not be dispositions that enable us to think for ourselves; they can also embed us in social contexts where we benefit from the thinking of others.

References

- Henrich, J. (2016). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton: Princeton University Press.
- Kihlstrom, J. (2013). The person-situation interaction. In D. Carlston (ed.) *The Oxford Handbook of Social Cognition* (pp. 786–806). Oxford: Oxford University Press.
- Skorburg, G. & Alfano, M. (2019). Psychological science and virtue epistemology: Intelligence as an interactionist virtue. In H. Battaly (ed.) *The Routledge Handbook of Virtue Epistemology* (pp. 433–445). New York and London: Routledge.
- Trout, J.D. (2005). Paternalism and cognitive bias. *Law and Philosophy* 24: 393–434.

4c Commentary from Quassim Cassam

Reply to Neil Levy

QUASSIM CASSAM

Levy thinks that we should very often be dogmatic when faced with what we recognize as possible evidence against our prior beliefs. We should simply ignore or refuse to consider such evidence when we have acquired our prior beliefs through testimony that is sufficiently good for those beliefs to be good candidates for knowledge. In such cases, “we ought to stand fast in the face of anomaly or evidence against our belief, unless the source of the evidence has the same kind of standing as the original source of testimony”. While arguing for dogmatism, Levy also objects to my account of the dispute between David Irving and Richard J. Evans about the reality of the Holocaust. Levy thinks that I overestimate the capacity of the ordinary intelligent person to adjudicate. Even if non-specialists read on Wikipedia that Irving was found by a British court to have distorted the historical evidence, this does not get them very far. For why, Levy asks, “should they accept either that Wikipedia accurately reports the court’s judgement or – more particularly – that the court was correct?”.

What are the epistemological and socio-political implications of the dogmatism that Levy recommends? Starting with the epistemological implications, I take it that for S to know that P, P must be true, S must believe that P, and S must have the right to believe that P. By simply ignoring what I recognize as possible evidence against my belief that P, I potentially deprive myself of the right to believe that P and thereby also potentially deprive myself of the knowledge that P. This is a high price to pay. However, Levy’s idea seems to be that I retain the right to believe that P, and hence my knowledge that P, as long as my prior belief that P came from a sufficiently good testimonial source. This is what permits me to ignore possible evidence against my prior belief without exposing myself to the charge that I no longer know that P. I can be dogmatic and still know that P.

Now consider the original testimonial sources of my beliefs about the Holocaust: books I read for school history lessons, what my

schoolteachers taught me, television documentaries and the odd movie or newspaper article. How do these sources compare with someone like Irving? Is their standing superior? That is *prima facie* unlikely. After all, Irving was a prolific author of books on historical subjects, mostly about the Second World War. Several of his early works were well regarded by fellow historians and some have been reprinted and reissued. It is true that he has no formal qualifications. However, as Evans remarks, ‘there are plenty of examples of reputable and successful historians whose lack of formal academic qualifications is as striking as Irving’s’.¹ I venture to suggest that Irving’s standing as a historian is superior to that of the schoolteachers, journalists and documentary film-makers whose testimony was the original source of my beliefs about the Holocaust.

In that case, I am *not* entitled to ignore Irving’s arguments by Levy’s own lights. If I am not swayed by those arguments, it is because Irving is not what Levy calls an “appropriately constituted epistemic authority”, but not because he lacks formal qualifications. The crux of the matter is that he was found by appropriate authorities – a British court, advised by Evans – to have deliberately distorted the historical evidence. However, when I base my rejection of Irving’s arguments on this fact, I am not being dogmatic. I have *reasons* for rejecting Irving’s views about the Holocaust *even after considering them*. I am not ignoring them. Someone might ask why I am so sure about Evans’ credentials or the reliability of reports about the court’s verdict. However, Levy faces similar questions: in defending a dogmatic response to Irving, he is making assumptions about Irving’s historical credentials in comparison to those of other sources of testimonial knowledge. These assumptions are no different from mine, and they are justified in the same way. It might be true, as Levy insists, that a layperson can only have a rough idea of the reasons for saying that Irving distorted the historical record but that is all the layperson needs for his or her rejection of Irving’s view to be non-dogmatic.

If this is correct, then Levy is not as dogmatic as he thinks he is. Having said that, it is worth reflecting on the socio-political implications of a policy of dogmatic non-engagement with evidence – even misleading evidence – against one’s prior beliefs. It is important for the citizens of democracy not only to know that certain claims are false but to have at least a very rough idea of *why* they are false. In the case of highly technical subjects like climate change, this might not be possible. However, the idea that the intelligent layperson is in no position to come to a reasoned conclusion about the relative merits of Evans and Irving as historians or, for that matter, the relative merits of 9/11 conspiracy theories and the official view is absurd. One might not know enough physics to be able to refute what conspiracy theorists say about the collapse of the Twin Towers but there are many other accessible reasons for rejecting 9/11 conspiracy theories after due consideration. The evidence of al-Qaeda’s

responsibility for 9/11 is overwhelming, and one doesn't need a degree in physics to know that.

When Holocaust deniers and other conspiracy theorists use their baseless speculations to manipulate public opinion and advance their political objectives it is important that ordinary citizens feel empowered to resist. It is not good enough to leave it to the experts, who probably have better things to do. It is vital that as many non-experts as possible understand and can explain to one another why the claims of Holocaust deniers are preposterous. To object that only experts are qualified to pronounce on these matters is to leave the field open to people like Irving to promote their ideas with no pushback from ordinary citizens. Some ideas are too toxic for responsible citizens to ignore, regardless of their academic qualifications. We must, as Kant insisted, have the courage to use our own understanding, even if that understanding is limited. The dogmatism that Levy favours is the antithesis of the enlightenment ideal of lay knowledge and understanding. It should be firmly rejected.

Note

- 1 Evans: David Irving, Hitler and Holocaust Denial | Holocaust Denial on Trial (hdot.org), Section 2.2.2.

4d Neil Levy's Response to Commentaries

Response to Commentaries

NEIL LEVY

I'm grateful to my two commentators for their very thought-provoking responses. Grateful but unmoved. Below, I set out my reasons for being, as at least one of my commentators would think, so unreasonable.

Quassim Cassam thinks "Levy is not as dogmatic as he thinks he is". In his eyes. I'm no more dogmatic than he is because I have *reasons* for assessing David Irving's credibility as low, consisting in the assessment of a British court. I share this reason with Cassam. I agree: I'm no more dogmatic than Cassam is, if he here reports his own reasons for disbelieving Irving accurately. But that's not because I'm not dogmatic; it's because he is.

Dogmatism, as it's used in these debates, consists in refusing to engage with the first-order evidence (for the purpose of making up one's own mind) for or against a claim. Of course, the court's judgement is not first-order evidence. To behave non-dogmatically in response to Irving would be to read his work and assess his arguments. I agree with Cassam that the court's verdict provides us with a reason for rejecting Irving. But I disagree that in taking that as my reason, I don't behave dogmatically.

Of course, we can use "dogmatism" however we like. Shorn of disputes over words, my claim is that we non-experts ought to rely on higher-order evidence (like the testimony of experts) and not first-order evidence. It's clear that my dispute with Cassam is not merely verbal: he's explicit that we ought not to leave these issues to the experts but assess them for ourselves. It's central to my argument that we can't in fact do this; our apparent non-dogmatic engagement is no such thing. Cassam rightly sides with Richard Evans against Irving, but in fact (I bet) he found Evans' arguments more convincing than Irving's because he was already disposed to defer to Evans. Detailed attention to the argument of Holocaust deniers, JFK conspiracy theorists, sophisticated anti-vaxxers and so on will reveal just how difficult it is for non-experts to assess such claims. Of course, that's an assertion (just as Cassam's claim that we can reliably and responsibly engage with such people is an assertion). Here's an empirical prediction to move the debate forward.

If we present the arguments of a sophisticated conspiracy theorist to a naïve (but conscientious and educated) audience, paired with arguments for the truth, experimental participants will do no better than chance at picking the correct view. Of course, actually testing this prediction will be difficult. We will need to identify a debate on which there is an expert consensus (so as not to beg any questions), but on which there is nevertheless sophisticated dissent of the conspiratorial sort, on a topic regarding which most people have no previously settled views. Perhaps the feasibility of perpetual motion machines or some topic in climate science apparently distant from the hot button issue (e.g., climate sensitivity) might play this role. Ideally, we would test a variety of such debates, across a variety of naïve audiences. If my prediction is correct, then engagement with the first-order evidence is not a reliable means of ascertaining the truth; by itself, this would constitute a powerful consideration in favour of my view.

Steven Bland is much more sympathetic to my project than Cassam is. He takes issue, however, with the suggestion that we ought to focus on environments *rather than* agents. Since behaviour and cognition is always the product of context and agent, there's no reason to think the former is a better focus of intervention, in general, or in principle. Take our disposition to defer to the prestigious. While this disposition has been epistemically beneficial in our evolutionary past, today it tends to mislead us, Bland suggests. We might respond *either* by changing the environment in which we operate as epistemic agents *or* by "cultivating greater epistemic discernment". Bland concedes there may be reasons to do the former rather than the latter, but these reasons are pragmatic rather than epistemic.

I deny, however, that the dispositions that constitute our epistemic vigilance (our conformity bias, prestige bias, our disposition to prefer testimony from the benevolent, and so on) are on all fours with the epistemic environment, such that we might in principle take them as just as appropriate targets for intervention. There are two reasons why we should prefer to change the environment. The first, shallow, reason is that we might find changes in our dispositions to defer difficultly. While the debate between theorists like Cecilia Heyes and more mainstream cultural evolutionists remains unsettled, it may be that these dispositions are robust to many environmental perturbations.

The second reason is stronger and more important to my overall view. It is this: despite the name of many of these dispositions (the prestige *bias*, and so on) these dispositions are constitutive of our rationality. In response to Cassam above, I noted that the heart of my view is that we ought to be guided by higher-order evidence when we lack the capacity to assess the first-order evidence. Cassam rejects this view as "the antithesis of the enlightenment ideal". I think this is a mistake. Higher-order evidence is genuine evidence, and in responding to it we respond

rationally. The prestige bias is rational because prestige is higher-order evidence that an agent's first-order beliefs are correct. Consensus is higher-order evidence; benevolence is higher-order evidence. And so on. We should leave our dispositions (more or less) as they are because they are constitutive of us as rational agents. Perhaps we could design agents who reliably form true beliefs on the basis of dispositions that are not responsive to genuine evidence *qua* evidence, but that project is one we will find difficult to undertake. The creatures that would emerge from such a design program would be radically different from those we take ourselves to be and which most of us want to be. We best use our understanding by ensuring the evidence the environment provides to us constitutes genuine reasons, rather than by building a creature that ensures reliability by perverse response to bad reasons.

T&F Proofs – Not for Distribution

5 Mindshaping and intellectual virtues

Alessandra Tanesini

Character is a human psychological feature that is not shared by other primates.¹ It is also the product of repeated activities whose function is primarily to shape the minds of those whom they target so that they acquire those settled global dispositions that constitute individual characters. Intellectual virtues are among the character traits that are brought into existence in this way. Hence, even though intellectual virtues are psychological traits of individuals, their acquisition and preservation are generally socially mediated. In addition, or so I argue in this chapter, the ultimate practical and epistemic ends that explain why human communities have shaped their members into creatures with virtuous character traits are inherently social. Human beings are constantly under social pressure to be intellectually virtuous because those with these traits are better able to coordinate their epistemic and practical activities with others in their community than those who lack these features. That is, individual virtues have been culturally selected for their social epistemological and practical benefits.

This chapter consists of four sections. The first introduces the notion of a mindshaping practice or activity and explains its role in cultural evolution. Ultimately, it is humans' evolving ability to shape each other's mind and susceptibility to having one's mind shaped that has enabled us to solve numerous coordination and mixed-motive problems thereby enhancing our ability jointly to perform practical and epistemic tasks.² Section 2 redescrines the processes and techniques of character building as examples of mindshaping and self-shaping whose primary function is to enhance mutual intelligibility in the service of solving coordination problems. It also shows that character attributions serve the purpose of making character rather than merely describing it. If successful, character attributions have the powers of self-fulfilling prophecies (cf., Alfano 2013). The third section focuses on intellectual virtues as the products and tools of mindshaping activities. The concluding section briefly sketches why this approach also promises to throw novel light on the evolution of intellectual vices.

1 Mindshaping

In its broadest sense, “mindshaping” refers to any activity whatsoever that leads some person to change some of her propositional attitudes, emotions, values or settled dispositions (Mameli 2001). For example, if I become angry as a result of being insulted, the insult is, in some sense, a mindshaping activity. Since most social exchanges are directly or indirectly concerned with changing the minds of those with whom one is interacting, mindshaping, in this broad sense, is both ubiquitous and highly heterogeneous.

In the philosophy of mind, mindshaping has emerged as an account of folk psychology that is an alternative to the traditional mindreading approach (McGeer 2007, 2015; Zawidzki 2013, 2018). Mindreading in either of its two main variants (theory-theory and simulation) holds that human beings are typically able to predict others’ behaviour by correctly figuring out the independently formed mental states that guide that behaviour. This figuring out is an epistemic task that is achieved either by theorising or by simulation. Irrespective of the mechanism, folk psychological attributions of beliefs, desires, emotions and character traits, according to these views, are empirical claims that correctly or incorrectly describe the mental states and traits of the persons one seeks to understand (Goldman 2006; Gopnik & Wellman 1994).

The mindshaping alternative holds instead that the primary function of folk-psychological attributions is to shape the target mind, so that it fits the attribution, rather than to describe that mind as it already is. For example, attributions of a belief to the self or to others would not aim to get right, or track, what the person already believes. Instead, attributions would function to get that person to form and sustain accurate beliefs. This account of folk psychological attributions, as McGeer (2015) points out, offers a natural explanation of the so-called transparency of belief. When asked whether they believe that *p*, in ordinary circumstances normal human beings do not answer by first introspecting the contents of their mind, instead they try to figure out whether *p* (Evans 1982). That is, in this case at least, a solicitation to engage in folk psychological belief self-attribution is treated as a request that one makes one’s mind up in accordance with the evidence, rather than as a solicitation to introspect.

Supporters of the mindshaping account of folk-psychological attributions extrapolate from this and other cases to argue that whenever people attribute propositional attitudes, emotions, character or personality traits to human beings, what they are doing (irrespective of their intentions) has the primary function of shaping the minds of those to whom these features are attributed. Hence, attributions of belief aim to get others to believe what is right (in accordance to the appropriate epistemic standards), attributions of desire to make them desire what

is proper (given some shared social norms) and in general to induce the targeted individuals to act in conformity with shared standards (McGeer 2015).

When mindshaping is advanced as an alternative to mindreading, its supporters need to show that mindshaping predates mindreading abilities, so that we can have the first without the second. If there is no evidence for such dissociation then mindshaping is best seen as a phenomenon that complements mindreading (Peters 2019), or one that essentially relies on the ability to mindread to get off the ground (Westra 2020). In this chapter, I am not trying to adjudicate this issue. For my purposes, it is sufficient that folk psychological attributions, and more specifically, trait attributions have a mindshaping function and that their prevalence and persistence are largely a function of their mindshaping powers. This empirical claim might be correct even though mindreading abilities are required for mindshaping to be effective.³

My focus in this chapter is on a broader range of mindshaping activities that comprises, but is not limited to, at least some folk psychological attributions. I do not, however, include any activity capable of causing a change in someone's mind. Instead, I restrict my attention to those activities whose proper function is mindshaping. These are actions and practices whose persistence and prevalence are due to their mindshaping powers.⁴ There are many uncontroversial examples of mindshaping so understood. For example, shaping minds is the explicit aim of teaching. The educator wishes her students to form new true beliefs as a result of her teaching. In addition, she might foster learning by creating a classroom environment (a cognitive niche) that scaffolds the students' studying so that they are better able to acquire novel true beliefs and understanding. I discuss a number of these practices in Section 2, where I argue that thinking of character as a product of mindshaping throws light on the role and value of character in human communities.

There are several different ways of classifying mindshaping activities. For instance, one may wish to focus on the mechanisms involved, such as imitation, social learning of a different sort, or individual learning in some social environment. Instead, I use here two different orthogonal taxonomic principles. The first distinguishes practices of self-shaping from activities where the mindshaper is distinct from the person whose mind is being shaped. Individual learning and exercises of self-control are examples of the first kind; explicit teaching and expressions of other directed negative reactive attitudes such as blame and anger of the second.

The second principle concerns the nature of the mindshaping intervention. My interest lies in two kinds. The first comprises activities that set normative expectations; the second of activities that express empirical expectations. Normative expectations include demands or requests that establish novel commitments or obligations.⁵ For instance, the

person who promises to herself that she will take a walk every day sets a new obligation for herself. The making of this promise is an activity of self-regulation. It is an act where one shapes one's own mind by creating a new reason to do something one might otherwise not be inclined to do. Ordinary practices of blaming, praising and rewarding people for their actions, emotions, traits and beliefs also aim to set, or reinforce, normative expectations designed to shape people's mind to conform to shared social norms and values.

Surprisingly, empirical expectations also have the power to shape minds. These expectations are predictions, rather than demands. Curiously, these can come true even when they are based on false assumptions. It is well known for instance that people who expect to recover from an illness, even when that expectation is not supported by the available evidence, have a better chance of recovery than those who form a realistic assessment of their prospects. The expectation of recovery causes one to feel optimistic, less stressed and more able to enjoy life. The psychological changes, in turn, impact the immune system and improve one's ability to fight infections. Expectations can thus become self-fulfilling prophecies (Snyder 1984). These effects of expectations that fulfil them are known as expectancy effects (Mameli 2001, 609).

These expectations can be self or other directed. For example, a person who thinks highly of her mathematical abilities also expects, in the sense of predicts, that she will perform well in a number of mathematical tasks. This expectation fills her with confidence which permits her to perform at the best of her abilities. The same confidence might also make her enjoy the challenge of solving mathematical problems. As a consequence, she practices doing mathematics, and her abilities improve. Self-directed expectations, which might have been poorly supported by the empirical evidence, causally contribute to bring about effects that confirm them.

By the same token, empirical expectations about other people can make them conform to the set expectations. For example, if parents expect their first born to inherit the family business, they might create an environment that facilitates in this offspring the acquisition of the skills required to lead a business. These parents might treat the second born differently by expecting him to follow his brother rather than to lead him. Since the two children find themselves in what are, in effect, different cognitive niches, they are likely to develop different behavioural dispositions which given the incentives set up by the parents are likely to make them best suited to the roles that the parents expected them to fulfil. This might occur without the second born ever thinking of himself as a follower. He might be put under less pressure by the parents who might also encourage interests unrelated to the family business.⁶

In several cases, however, the internalisation of the expectations in the target's self-conception plays a causal role in the generation of the

expectancy effects. A paradigmatic example of this phenomenon is the transmission of gender stereotypes. Adults' different gendered expectations of children's behaviours lead them to treat male and female babies very differently from each other. Because they find themselves in different social environments, children acquire different behavioural tendencies depending on their gender. These differences in dispositions are sharpened as children learn to identify with the gender attributed to them. Subsequently, children internalise gender stereotypes in their self-conceptions. As a consequence, they believe the stereotypes and act them out (Snyder & Klein 2005).⁷

These examples give an initial flavour of the heterogeneous practices and activities that shape human minds. Some of these are deliberately performed to this end as they consist in the creation of normative expectations to change opinions and behavioural dispositions. Others have these effects, although actors do not always intentionally set out to achieve this end. It is plausible that humans' heightened receptivity to being shaped by conspecifics and tendency to engage in activities that result in the shaping of minds are the result of cultural evolutionary pressures. That is to say, even without genetic mutations of any sort, some human beings living in communities have acquired novel abilities that can be transmitted horizontally and vertically across generations. Because these abilities give an advantage to those who possess them, overtime they become more and more prevalent in the population.

I borrow an example from Mameli (2001) to illustrate the point. Suppose an early human, for whatever reason, behaves toward her babies in ways that imply that she attributes to them a precocious ability to communicate. Unlike her contemporaries, she treats the babies' non-sense vocalisations as attempts to communicate with her. Hence, whenever a baby vocalises, she rushes toward the child or pays her special attention. The baby thus learns to associate these sounds with her mother's appearance and begins to use the sound as a call. Thanks to their mother's repeated communicative engagement with her, this baby, and her siblings, learn to speak earlier than other children and also acquire superior communicative abilities. These abilities give them an edge in their community. Further, when they have children themselves, they adopt their mother's child-raising practices. Thus, these superior abilities are transmitted down the generations. Further, other members of the community might also adopt the same practices having observed the success of children raised that way. Of course, once these novel abilities are entrenched, they might make further mindshaping practices possible and thus generate cascading effects.

More generally, work on human cultural evolution strongly suggests that human social cognitive abilities have evolved in the direction of more and more refined communicative abilities in the service of finding

better solutions to the coordination problems encountered by early humans (Sterelny 2012). This is because coordination is made easier if all adhere to the same coordination-facilitating norms which include communicative conventions. Further, when tasks are complex and better handled by experts, coordination is enhanced if individuals specialise in different activities. Humans would have solved these problems by adopting divisions of labour in accordance with clearly visible markers (O'Connor 2019).

Mindshaping activities are uniquely well suited to facilitate the kind of maximal mutual intelligibility instrumental to solving coordination problems. It is much easier to coordinate one's actions with a person who conforms to what we expect of them than with one who does not. The effect of mindshaping is to bring oneself and others to behave in accordance with the same norms and to be intrinsically motivated to follow them. The result is the creation of what McGeer (2015) calls practice-dependent epistemic advantages. The person whose actions are regulated by some norms is in a better position to understand and coordinate with another who plays by the same rules, than with any person who follows different ones.

2 Shaping up: acquiring and retaining character

Character is the product of mindshaping activities. I take this claim to be a near platitude. In this section, I first highlight some character-forming practices and show that they are examples of mindshaping. These practices include, but are not limited to, character trait attributions whose main function is to steer people toward virtues and away from vices. Thinking of character as a product of mindshaping shows that character matters primarily because it enables success in joint practical and epistemic activities which require coordination. It is not only intellectual virtues that are advantageous in this way, all other character traits including moral virtues facilitate coordination since what matters for this purpose is that individuals' dispositions are stable and cross-situationally consistent.

It is the main contention of this section that shaping agents to have characters offers advantages in the service of coordination that are additional to moulding individuals into following the norms characteristic of beliefs and desires. McGeer (2015) highlights that mindshaping practices direct people to follow the same norms as each other. These practices would thus be analogous to training everyone to play by the rules of chess rather than, say, drafts. Here I argue that those practices that specifically mould individuals to acquire and retain character traits are instrumental to training people to play the game continuously, rather than only engaging sporadically. Thus, any character trait, including intellectual virtues, makes its possessor more intelligible to others who

play the same game, and thus facilitates coordination. In the next section, I argue that intellectual virtues specifically generate further advantages in the pursuit of joint activities.

Personality traits are global dispositions to behave, feel and think in specific ways when the circumstances are relevant. Hence, neatness, extroversion, courage, stinginess, rudeness and closed-mindedness are all personality traits. The ascription of these features to individuals highlights that it is expected that they will behave in accordance with the trait over time and across a range of different circumstances. Hence, personality traits are dispositions that are both stable and cross-situationally consistent.⁸

Arguably there is an ordinary understanding of character that admits that all personality traits are part of character. In this sense, someone's character is her personality. For my purposes here I adopt a narrower notion of character that identifies character traits as a proper subset of personality traits. According to this view, only those personality traits for which people are normatively evaluated are part of their character and thus correctly identified as character traits (Miller 2014, 15). Hence, closed-mindedness, courage and rudeness would be character traits, but extroversion would be best thought of as a mere personality trait.⁹ Character traits would thus include moral and intellectual virtues and vices.

Even a moment's reflection reveals that all the strategies that aim to form characters are examples of mindshaping in the broad sense of being activities whose function is to produce in those targeted by these strategies novel settled dispositions that match a model. These strategies are predicated on the assumption that character is acquired, and that it can be moulded. Further, the strategies are consciously adopted precisely because of their alleged efficacy in shaping minds so that they exemplify those model character traits which the mindshapers wish to inculcate in others.

These character-forming strategies include:

Explicit Teaching. In some settings, including but not restricted to formal education, young people and adults are told that some character traits labelled as virtues are worth pursuing for their own sake. Teaching creates incentives to behave in accordance with the virtues such as rewards for behaviour that is consonant with these traits. In addition, when teachers expect, in the sense of predict, students' compliance, these expectations themselves might cause students to behave in ways that fulfil them. Students, for instance, might get a sense of satisfaction from meeting the standards set by teachers.

Exposure to Exemplars. Adults and children meet in real life, and are presented with narratives about, people whose characters are held as exemplary by those who surround them. The emulation of exemplars probably starts as imitation early in life. Parents seek to be examples for their children. Older siblings are also often told to set an example. These

practices explicitly rely on children's imitative propensities to shape their minds towards the acquisition of virtue. Many tales and novels for children also involve exposure to characters that are hailed as exemplars to imitate or to shun. These narratives often seek to inculcate in children the belief that virtue is rewarded (or that it is its own reward), while vice is punished.

Individual learning in cognitive niches. People acquire virtuous dispositions by practice. Even though this is done by individual learning, others can engineer the social environments that promote such practice. Hence, for example, parents who want their children to be courageous might put them in controlled situations that force the children to face danger. Even though the children learn to be courageous by learning to control their fears, their brave behaviour is also an expectancy effect of the parents' expectations. In turn, repeated courageous actions, together with the satisfaction of meeting parents' expectations, facilitate in the children the acquisition of those settled dispositions that are characteristic of courage.

Undoubtedly these characterisations are far too brief but they should suffice to indicate that the most common strategies of character formation are techniques to shape minds that rely on explicit rewards and punishments (normative expectations) and/or empirical expectations to bring minds to fit what is expected (normatively and empirically) of them. While this conclusion should be, on reflection, quite obvious, it is certainly less obvious that the practice of attributing character and personality traits to individuals is also best thought as an instance of mindshaping. Yet personality and character trait attributions offer a clearer example of mindshaping than the ascription of propositional attitudes considered by some supporters of the mindshaping account of folk-psychology (cf., McGeer 2015).

The mindshaping features of character trait attributions are at its most transparent when we consider the explicitly evaluative nature of virtue- and vice-ascriptions. In most contexts to say of people that they are open-minded, courageous or generous is a way of praising them. By the same token, to claim that someone is closed-minded or cowardly or stingy is to disapprove of them. We use this vocabulary as a way of enjoining people to preserve and develop further whatever virtues we attribute to them, and to change so as to eliminate or at least lessen whichever vicious features we ascribe to them. If this is right, then the ascription of character traits is at least in part a practice whose aim is to strengthen virtue and weaken vice.

Folk psychological character trait attributions are thus a component of the practice of responding to each other by expressing a range of negative and positive attitudes like anger, guilt, hurt feelings, gratitude or admiration. Expressions of these attitudes convey normative expectations and thus supply reasons but also incentives to shape one's behaviour and

mind so that it fits whatever is classified as praiseworthy or admirable in accordance with shared practices, and avoids that which is disapproved. In short, folk-psychological character attributions wear on their sleeve their evaluative nature as expressions of normative expectations that purport to influence minds and actions. For this reason, they – and the reactive attitudes with which they are closely connected – are best thought of as contributing to mindshaping practices.

These folk-psychological ascriptions also have the self-fulfilling power of some empirical expectations. For instance, Richard Miller and colleagues (1975) have shown that telling students that they are tidy made them become tidier than a control group but also neater than those who were exposed to arguments in favour of tidiness. It appears that the students incorporated the label into their self-conception, thus becoming the tidier persons that they thought the experimenters took them to be. By the same token people who become aware of stereotypical attributions might subsequently acquire the traits that conform to the stereotype. For example, young girls learn very early on that girls are supposed to be more fearful and less aggressive than boys. In response girls often become less courageous and more docile than boys, they do so partly in reaction to how adults relate to them, partly through internalising the attributions about them made by adults (Klein & Snyder 2003).

Alfano (2013) has offered a detailed account of how folk-psychological character attributions can function as self-fulfilling prophecies so that those who are labelled virtuous frequently change their behaviour but also motivations and thoughts to fit the label applied to them. Whilst my analysis is largely in agreement with Alfano's, I wish to take issue with two aspects of his view. First, Alfano interprets virtue labelling as a kind of mindreading that, whilst false, has the additional power to bring about its own truth (2013, 106). Such labelling is thus something akin to a noble lie that turns fiction into fact. This is why factitious virtue would be factitious. Second, Alfano claims that factitious virtue is always motivationally distinct from ordinary virtues because the person whose virtue is factitious is 'in part motivated by a desire to maintain his self-concept' (2013, 101). That is, the expectancy effects of factitious virtue would always be mediated by incorporation into the self-concept. Hence, factitious virtue would only simulate real virtues without being identical to them.

With regard to the first point, Alfano resorts to claiming that virtue labelling is an indirect speech act where one uses an assertion to make a recommendation (2013, 106). In his view, virtue attributions, in addition to expressing normative expectations, would involve false claims about people's psychologies. In my account, instead, virtue labelling is a prediction that, because it is made, creates new incentives for its target to act in accordance with it. So understood, virtues would be factitious in the sense of being something that is partly manufactured through

being ascribed. It is not fictitious, however, since trait attributions are not false assertions about independently existing psychological features of the target.

My disagreement with Alfano on the second point goes deeper. Alfano seems to think that the person whose virtue is factitious is ultimately partly motivated by the need to maintain a positive conception of the self. Since this motivation is not wholly virtuous, factitious virtue would only simulate the real thing, but be distinct from it. I think he is in this regard mistaken. Alfano's mistake in my opinion lies in isolating virtue labelling from other forms of mindshaping activities with which it is connected.

Virtue labelling is only one of the many practices that have evolved to shape human minds and behaviours. These practices do not create individuals whose good motivations are actually dependent on others' approval in the service of self-esteem. Instead, they produce genuine virtue because they create minds that are disposed to act virtuously out of virtuous motivation.¹⁰ I shall return to this point below when I discuss the role of intellectual virtues in promoting cooperation among cognitively diverse agents.

There is, however, at least one respect in which the products of mindshaping differ from virtuous traits as these are traditionally understood. The former but not the latter require continuous scaffolding and support. That is, techniques of mindshaping must operate continually to sustain a match between agents' attitudes and dispositions and the model or standards to which they are normatively and empirically expected to conform. When these scaffolds are removed, we should expect overtime agents to fall out of step with shared models. In short, mindshaped character traits are rendered stable by the continuous presence of external (and internal) scaffolds. Whilst this is a difference with virtue as traditionally conceived, the latter would also require continued application to be sustained. Further, mindshaping includes self-regulation in the form of undertaking commitments. Hence, the importance of this point of difference should not be overestimated.

I have argued so far that character is the product of mindshaping activities that set normative and empirical expectations and that are often deliberately designed to bring about mindshaping effects. Thinking of character as a product of mindshaping makes sense, once we notice that the maximisation of mutual intelligibility as a means to achieving coordination is the proper function of mindshaping. People who have characters, as well as beliefs and desires, have dispositions that are diachronically stable and cross-situationally consistent. Character traits would thus be internal scaffolds that help to stabilise one's behaviour over time and in different circumstances. Coordinating activities with people who have these traits is much easier than coordination with rational agents who are very susceptible to situational factors.

For example, compare two agents both of whom tend to regulate their beliefs in accordance with the evidence in their possession. These agents would usually have the same doxastic attitude about whether p , provided that they have the same evidence in their possession. Both agents are intelligible to someone with that evidence and who regulates her beliefs by the same evidential rules. Suppose, however, that one of these two agents is diligent, while the other is frequently apathetic.¹¹ The first individual always believes in accordance with the evidence in her possession which she carefully assesses. The second agent's behaviour is more erratic. On some occasions, he forms beliefs in accordance with his current evidence, but on others he is careless. Thus, these two agents often form different beliefs because only the first is able assiduously to follow the norms of belief. Coordinating activities among agents who are diligent is easier than coordination among agents who are idiosyncratically apathetic, or among groups including both kinds of agents. Agents of the first kind are more likely to be in step with each other over time and across situations than agents of the second kind. This is because character traits make agents' behaviour more stable and thus more intelligible.¹²

What I have said for diligence is also applicable to other character traits including those that are not virtues. The acquisition of character makes one's behaviour more regular, less susceptible to situational factors that are not controllable such as the weather. In this regard, even vice is preferable to characterlessness. Whilst vicious persons cannot be relied on if they are dishonest or lazy, it is possible at least to rely on the stability of their vices. So although mindshaping practices serve the function of moulding minds that among other things exhibit virtuous characters, persons whose minds have been shaped into vice are still more intelligible than people of no character.

3 Intellectual virtues and mindshaping

I have argued that human beings that possess character traits in addition to beliefs, desires and other propositional attitudes are more likely to coordinate their activities successfully because they are mutually more intelligible than those who lack these traits. The advantage conferred by the possession of character is the result of the stability and cross-situational consistency of those behavioural dispositions that are an essential aspect of character. This stability and consistency facilitate coordination especially when all participants in an activity share the same character traits so that their propositional attitudes and actions would normally be expected to be largely in sync with each other.

These considerations do not take into account that many human practical epistemic activities are carried out more successfully by groups that divide cognitive labour among participants who specialise in performing

different tasks. The institution of division of labour has two important consequences with regard to subjects' ability to coordinate and willingness to cooperate. First, the development of specialisation brings cognitive heterogeneity in its trail. Second, specialisation makes it easier for some agents to free ride on others' labour. In this section, I argue that the possession of intellectual virtues is crucial in turning situations where there are conflicts of interest into coordination problems because they supply the intrinsic prosocial motivations necessary to avoid free-riding.¹³ Often, these are coordination problems that are best solved when actors adopt complementary strategies, rather than act in the same manner (O'Connor 2019, 31–33). In this regard, the intellectual virtues confer advantages additional to those conferred by the cultural evolution of other non-virtuous character traits.

The best way to address some problems, especially those whose solution requires possession of sophisticated skills, is to divide labour among group members. Different individuals are trained to perform different tasks so that together they are able to achieve their goals more reliably and efficiently. Such division of labour has proved effective to solve practical problems but also to carry out inquiries. For this reason, the vast majority of scientific research is performed by teams where individuals are allocated different tasks, and where junior members are often trained to acquire some specific skills. Hence, research specialisation is a source of cognitive heterogeneity.

The promotion of different skills in different subsets of the population creates opportunities for free-riding. There are situations in which the best outcome for each individual is to gain from others' labour without contributing a fair share. Of course, if all act in this non-cooperative manner, they all lose out. But if one manages to deceive one's partners then one gains from their labour without having to expend energy. In situations in which all joint activities are carried out together, and information is shared publicly among all members, publicness by itself is an obstacle to free-riding (Sterelny, 2012, ch. 5). But when a group specialises, some activities are carried out by some individuals alone or in subgroups. In these contexts, free-riding can be pulled off more easily because one can hide one's activities from public scrutiny (Zawidzki 2013, 102–103).

Mindshaping individuals into acquiring, and retaining, intellectual virtues offer a solution to the challenges posed by cognitive heterogeneity and increased opportunities for free-riding that are the necessary by-products of division of cognitive labour and specialisation. I have characterised virtues, including intellectual virtues, as comprising those character traits for which individuals are admired, and which they are encouraged to achieve. Two features of intellectual virtues single them out as solving these two obstacles to coordination in conditions in which individuals would have incentives not to cooperate. First, intellectual

virtues promote cooperation among members of the same group because they supply the necessary pro-social intrinsic motivations. Second, they make epistemic dependence on cognitively diverse individuals mutually beneficial.

Intellectual virtues, and virtues in general, are character traits for which individuals are normatively evaluated. Further, these traits involve intrinsic motivations to act in accordance with virtue. Hence, for example, open-mindedness requires that one engages appropriately with alternative viewpoints out of a love for epistemic goods such as knowledge and understanding (Baehr 2011a). The intrinsic epistemic motivations characteristic of intellectual virtues are in effect prosocial motivations that promote cooperation.¹⁴ The person who acts open-mindedly out of an intrinsic concern for the truth is not likely to subordinate evaluating fairly views that are alternative to her own to gaining a personal advantage.

Whilst intrinsic epistemic motivations are in general pro-social, there are also virtues whose motivations are explicitly concerned with others' access to epistemic goods. These are the so-called virtues of epistemic dependability (Byerly 2021). They include epistemic benevolence, sincerity, communicative clarity and the virtues of offering good epistemic guidance to those whom one is teaching. These are those virtues that make an agent ideally suited to being the kind of person upon whom others can depend to gain knowledge and understanding and to acquire or maintain epistemic abilities and skills.

The practices that shape individuals to acquire and retain intellectual virtues are practices that lead those who have been shaped to see some norms as intrinsically motivating so that they are prepared to follow them even when compliance is costly. These practices include explicit teaching of the norms but also presentations of idealised exemplars by way of fables and other narratives. They also comprise systems designed to enforce compliance with norms by punishing counter-normative behaviours. Those who are intrinsically motivated to be intellectually virtuous are less likely to free ride and are, instead, disposed to cooperate.

Surprisingly, humans are also intrinsically motivated to punish since they are prepared to sanction others even when doing so is to the detriment of the punisher. For instance, cross-cultural studies have shown that human beings tend to be reciprocators. In Ultimatum games they are prepared to take home nothing in order to punish those who offer them little (Henrich & Henrich 2007).¹⁵ In addition, human agents deploy forms of self-regulation to commit to desires whose realisation would require costly activities. This is a way of transforming a mere desire into a value, and potentially into a goal which one is intrinsically motivated to pursue (cf., McGeer 2015, 264). All of these techniques are forms of mindshaping that promote compliance with shared norms that one is intrinsically motivated to follow.¹⁶

There is some evidence that traits with the intrinsic motivations characteristic of intellectual virtues are the product of mindshaping and have culturally evolved because they promote coordination among cognitively heterogeneous individuals (Zawidzki 2013, ch. 4). I cannot fully defend this empirical claim here but some recent empirical results about the associations between intellectual humility, perceptions of dissimilarity and prejudice are suggestive in this regard. People who measure high in intellectual humility are less prejudiced than those who are less humble against people with whom they disagree. However, intellectually humble persons are also more inclined to trust selectively and to be distrustful of those whom they judge not to be humble (Alfano & Sullivan 2021; Colombo et al. 2020). This intellectual virtue would, thus, combine a propensity to open-mindedness within an in-group and a sceptical attitude to people perceived as members of an out-group. This combination of dispositions makes sense if intellectual humility has been selected because it facilitates cooperation within a conformist group that is also cognitively heterogeneous.

I have argued that intellectual virtues, because of their intrinsic motivations, promote cooperation even among agents that have some degree of cognitive heterogeneity and that operate in conditions where opportunities for defection are present. In what follows I explore how intellectual virtues create the conditions in which epistemic dependence, which is an inevitable consequence of specialisation, is largely mutually beneficial.

In order to make this point it is helpful to group virtues into three categories that are not mutually exclusive and might not be exhaustive. The first comprises those intellectual virtues that contribute to carrying out inquiries in an epistemically responsible manner. These include, for instance, inquisitiveness and open-mindedness. The second category is that of the virtues of epistemic dependability which I have introduced above. The third category consists of those intellectual virtues that make one the sort of person who is just in their epistemic transactions with those upon whom one might epistemically depend. These virtues will include testimonial justice (Fricker 2007); the virtues characteristic of good listeners and those who exhibit proper trust in relation to expertise (Zagzebski 2012).

Intellectual virtues belonging to the first category promote conformism among inquirers that are cognitively heterogeneous because they have differing roles, interests, capabilities and levels of skill. Open-mindedness, for instance, is promoted for novices and experts alike. It is admired in anyone irrespective of context and social role. Such uniformity of motivation, if achieved, would promote the kind of mutual intelligibility that makes coordination easier. Of course, most people often are not very open-minded. Nevertheless, mindshaping practices are effective at making people more open-minded than they would otherwise be. In this way agents who otherwise have different capabilities and

information are more intelligible to each other and thus more capable of coordination than they would if they had no character traits or if their traits were wholly heterogeneous. These epistemic and practical advantages brought about by intellectual virtues are independent of their role in promoting the acquisition of knowledge and understanding in inquiry. Even in cases where open-mindedness might lead one astray from the truth, possessing this trait makes one better able to understand others and to be understood by them (provided that they are also open-minded).

Intellectual virtues in the second category include motivations to promote the acquisition and retention of epistemic goods and cognitive skills in other people. They are thus characteristic of those who can be depended on not to exploit others' vulnerability to deception and misinformation. These are virtues that contribute to trustworthiness because they motivate people to treat other agents' normative expectations of assistance with their epistemic needs as reasons to assist. That is, epistemically dependable people take others' requests for help as reasons to help. For example, the person who is communicatively clear is motivated to communicate clearly because others' normative expectations that she communicates clearly are a reason for her to communicate clearly.

The virtues of epistemic dependability are, thus, the virtues of trustworthiness in the affective sense that others' trust in one is taken by one to be a reason to fulfil their normative expectations (Faulkner 2014). The acquisition by every agent of these virtues improves communication since no one who has these traits withholds information needed by others that is in one's possession. Enhanced communication thus facilitates coordination, among individuals who, because they carry out distinct tasks in the context of joint epistemic activities, are likely to have access to different bodies of knowledge. In addition, the virtues of epistemic dependability, when combined with the virtues of responsible inquiry, motivate individuals to take up the role of teacher or educator. The practices designed to instil dependability in all students and apprentices also prepare them for their future roles as educators of the subsequent generation.

Intellectual virtues in the third category include motivations to relate appropriately to those upon whom one depends epistemically. Hence, these virtues are characteristic of those who adopt a trusting attitude towards other agents. This kind of trust is not mere reliance but involves the normative expectation that others will do as we trust them to do precisely because of the trust that we invest in them (Faulkner 2014). The acquisition of these virtues in every agent contributes to better lines of communication since they promote the acquisition from others of knowledge that one needs but does not have. In addition, these virtues when combined with those of responsible inquiry, motivate people to take up the role of student or apprentice. The ability, and willingness, of humans to learn from each other clearly contributes to solving jointly problems through sharing information.

If the considerations offered here are on the right track the acquisition and preservation of intellectual virtues should be seen as the product of mindshaping practices. Humans teach, cajole, encourage and incentivise each other to develop these traits because possessing them has distinctive epistemic advantages for the community of inquiry. In the context of complex problems whose solution requires specialisation and its attended cognitive heterogeneity, virtues provide the motivations required to avoid free-riding, the degree of conformism necessary for mutual intelligibility, but also the motivations to assist others' overcome their epistemic vulnerabilities and to accept help with one's limitations.¹⁷

The discussion so far has focuses on intellectual virtues as the product of mindshaping, but it also suggests that these same virtues are also tools by means of which humans shape theirs and others' minds. I conclude this section with two examples of intellectual virtues that are themselves instruments of mindshaping: propaedeutic trust and the virtues of the will.

Adults and teachers sometimes trust teenagers, children or students to do something, even though they do not confidently predict that those in whom they put their trust will act as they are trusted to do. By adopting a trusting attitude adults set up normative expectations for their charges to live up to. The setting of these normative expectations is an example of a mindshaping practice that is effective because it creates a new incentive to act as expected if one wants to avoid the costs associated with disappointing those who have some power over one. But the institution of a novel normative expectation also creates a new reason to fulfil the expectation. Provided that the recipient of trust has already acquired some dispositions to be trustworthy, the trusting person by expressing trust makes themselves vulnerable to those whom she trusts. The creation of this novel vulnerability supplies the recipient of the trusting attitude with a novel reason to do as they are trusted. In this way, the virtue of trust is a mindshaping tool that moulds others into matching more closely the virtues of trustworthiness.¹⁸

The so-called virtues of the will include perseverance, diligence and self-control among others (Roberts 1984). These are the moral and intellectual virtues of will power. These virtues are forms of self-regulation that enable one to shape one's mind into committing to sustaining valued behaviours and attitudes. So conceived the virtues of willpower are the dispositions that enable the development of more sophisticated practices of shaping one's mind to match norms that one implicitly or explicitly endorses. These intellectual virtues would thus play an auxiliary role whose primary function is to facilitate the acquisition and maintenance of the other virtues by shaping and keeping one's mind in the shapes characteristic of these other virtues.

4 Concluding remarks

This chapter has demonstrated that intellectual virtue is a product and tool of mindshaping practices in the service of joint epistemic activities that has culturally evolved because it maximises mutual intelligibility and facilitates cooperation. Hence, even though intellectual virtues are individual character traits, their genesis, function and functioning are wholly social. Virtues are acquired as a result of mindshaping practices that are social in nature. These traits have culturally evolved to facilitate coordination in the context of social divisions of cognitive labour. In addition, they are sustained through the continuing operation of empirical and normative expectations that scaffold minds to retain virtuous dispositions and motivations.

Even though I lack the space to address this issue here, the framework that I have presented in this chapter also promises to throw light on the socio genesis of at least some intellectual vices such as intellectual arrogance and servility which are distortions of the virtues of trustworthiness and trust. Arrogant individuals are not disposed to respond appropriately to others' epistemic vulnerabilities, those who are servile have adopted deferential attitudes that make them extremely vulnerable. Intellectual vices such as these might be interpreted as the product of mindshaping strategies that promote success in joint epistemic activities while unfairly distributing the benefits of this success among the participants. It also raises the possibility that other vices might instead be maladaptations that have also emerged from these unfair distributions.

It is often noted that inequities can emerge when divisions of labour are pegged to visible social identities. Coordination is easier if tasks are divided by easily identifiable groups. But such divisions might also mean that some groups gain more than others from the collective successes. The gendered nature of several putative intellectual virtues and vices including intellectual humility and modesty, timidity, servility and arrogance suggests that something of this sort might be at play in the emergence of intellectual vices.

Acknowledgements

Thanks to Mark Alfano for his comments on an earlier draft.

Notes

- 1 Individual non-human animals might be different in temperament from each other so that some take more risks than others for example. However, we typically do not think of some individual non humans as more courageous, or more open-minded than others.
- 2 A coordination problem occurs when there are no conflicts of interests in so far as all involved wish to coordinate their activities in order to succeed.

- A mixed-motive problem occurs when cooperation is costly for at least some of the actors involved (Bicchieri 2006, 2–3).
- 3 Mameli (2001), for example, presumes that some mindshaping presupposes mindreading abilities.
 - 4 Zawidzki (2013, 2018) defines mindshaping as any cognitive mechanism whose proper function is to make a mind match a (behavioural) model by shaping it to acquire dispositions to behave like the model. Roughly speaking, the proper function of a mechanism is what that mechanism has been selected for (2018, 31).
 - 5 Normative expectations are expectations that license normative statuses. They can serve to bring these statuses into existence as is done in promising, requesting or ordering. They can function to support these statuses by censoring behaviours that contravene them and rewarding compliance. They can also serve to affirm the presence of these statuses. I thus use the term differently from Bicchieri for whom normative expectations are second order beliefs about what others believe should or should not be done, believed and so forth (Bicchieri 2017, 69, n. 10).
 - 6 Explicitly wanting to fulfil the parents' expectations so as not to disappoint them often also plays a role. The mere existence of the expectations is thus also an incentive.
 - 7 As this final example illustrates often normative and empirical expectations combine to supply both reasons and incentives to conform with expectations. Further, often these reasons and incentives are the result of societal expectations, rather than those of single individuals (cf., Bicchieri 2017).
 - 8 I largely set aside here situationist worries about the existence of these traits. Recently, the robustness of the results on which these worries are based has also been called into question (Alfano 2018).
 - 9 But note that there is a tendency to evaluate people even for their extroversion or their neatness. Hence, people attribute moral overtones to dispositions to be tidy or messy.
 - 10 In my view the normative expectations which are adhered to by those who act virtuously are discretionary rather than mandatory. Mere failure to meet these expectations results in disappointment rather than in the kind of disapproval that is meted to those who stray into vice.
 - 11 The same point could be made for moral character since coordination among the brave is easier than coordination among those who are on occasion brave but sometimes cowardly. Note that any cross-situational consistency in dispositions facilitates coordination since a group of cowardly individuals also know what to expect of each other.
 - 12 I am presupposing here that character traits tend to have high fidelity and thus admit of very few exceptions. See Alfano (2013) for the notion of high-fidelity virtue.
 - 13 In Bicchieri's (2006) social norms play this role by transforming mixed-motive games into mere coordination problems.
 - 14 In this chapter I presume rather than defend the view shared by several epistemologists that virtues comprise intrinsic epistemic motivations (cf., Baehr 2011b; Byerly 2021; Zagzebski 1996). I take the plausibility of the view that virtues are the product of mindshaping to add further plausibility to this view since the creation of intrinsic motivation is a major feature of mindshaping.
 - 15 Ultimatum games are one-shot interactions between strangers where one player offers a proportion of a fixed sum to the other player. If the second player accepts the offer, the first player keeps the whole sum minus what she has offered to the other player who keeps what he has accepted. If the second

- player rejects, both get nothing. In this context, it would be rational for the first player to offer as little as possible to the second who rationally should accept any offer not matter how small. This is not how humans usually behave in these circumstances.
- 16 These norms are instituted by normative expectations. It is a mistake in my view to think that these norms are in every case mandatory obligations. Instead, some normative expectations are discretionary obligations. These supply reasons to do something and warrant disappointment if they are not complied with. They do not however license the kind of reactive attitudes that are warranted by not doing what one is mandated to do. For example, orders institute mandatory obligations while requests create discretionary ones. Failure to comply with either warrants different responses. In my (2020) I discuss the role of discretionary obligations in testimony.
- 17 Levy and Alfano (2020) have derived very different lessons about individual intellectual virtues and vices from our best theories of human cultural evolution. They argue that cumulative cultural knowledge requires passive imitation on the part of individual agents. They also think that conformism despite its knowledge producing effectiveness is best thought as individual vice. Instead, I wish to highlight the plurality of mindshaping mechanisms and the intellectual virtuousness of adopting a trusting attitude. This plurality also show that mindshaping is not mere indoctrination since it can contribute to scaffolding the rational agency of its targets.
- 18 On hope and propaedeutic trust as a mindshaping instrument see McGeer (2008).

References

- Alfano, M. (2013). *Character as Moral Fiction*. Cambridge: Cambridge University Press.
- Alfano, M. (2018). A Plague on Both Your Houses: Virtue Theory after Situationism and Repligate. *Teoria*, 38(2), 115–122. doi:10.4454/teoria.v38i2.
- Alfano, M., & Sullivan, E. (2021). Humility in Social Networks. In M. Alfano, M. P. Lynch, & A. Tanesini (Eds.), *The Routledge Handbook on the Philosophy of Humility* (pp. 484–493). London: Routledge.
- Baehr, J. (2011a). The Structure of Open-Mindedness. *Canadian Journal of Philosophy*, 41(2), 191–213. doi:10.1353/cjp.2011.0010.
- Baehr, J. S. (2011b). *The Inquiring Mind: On Intellectual Virtues & Virtue Epistemology*. Oxford: Oxford University Press.
- Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.
- Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. New York: Oxford University Press.
- Byerly, T. R. (2021). *Intellectual Dependability: A Virtue Theory of the Epistemic and Educational Ideal*. London: Routledge.
- Colombo, M., Strangmann, K., Houkes, L., Kostadinova, Z., & Brandt, M. J. (2020). Intellectually Humble, But Prejudiced People. A Paradox of Intellectual Virtue. *Review of Philosophy and Psychology*, 12, 353–371. doi:10.1007/s13164-020-00496-4.
- Evans, G. (1982). *The Varieties of Reference*. Oxford: Clarendon Press; Oxford University Press.

- Faulkner, P. (2014). *Knowledge on Trust*. New York: Oxford University Press.
- Fricker, M. (2007). *Epistemic Injustice: Power & the Ethics of Knowing*. Oxford: Clarendon.
- Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford and New York: Oxford University Press.
- Gopnik, A., & Wellman, H. (1994). The Theory Theory. In L. A. Hirschfield & S. U. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 257–293). New York: Cambridge University Press.
- Henrich, N., & Henrich, J. P. (2007). *Why Humans Cooperate: A Cultural and Evolutionary Explanation*. Oxford and New York: Oxford University Press.
- Klein, O., & Snyder, M. (2003). Stereotypes and Behavioral Confirmation: From Interpersonal to Intergroup Perspectives. *Advances in Experimental Social Psychology*, 35, 153–234. doi:10.1016/S0065-2601(03)01003-7.
- Levy, N., & Alfano, M. (2020). Knowledge from Vice: Deeply Social Epistemology. *Mind*, 129(515), 887–915. doi:10.1093/mind/fzz017.
- Mameli, M. (2001). Mindreading, Mindshaping, and Evolution. *Biology & Philosophy*, 16(5), 595–626. doi:10.1023/a:1012203830990.
- McGeer, V. (2007). The Regulative Dimension of Folk Psychology. In D. D. Hutto & M. M. Ratcliffe (Eds.), *Folk Psychology Re-Assessed* (pp. 137–156). New York: Springer-Verlag.
- McGeer, V. (2008). Trust, Hope, and Empowerment. *Australasian Journal of Philosophy*, 86(2), 237–254.
- McGeer, V. (2015). Mind-Making Practices: The Social Infrastructure of Self-Knowing Agency and Responsibility. *Philosophical Explorations*, 18(2), 259–281. doi:10.1080/13869795.2015.1032331.
- Miller, C. B. (2014). *Character and Moral Psychology*. Oxford: Oxford University Press.
- Miller, R. L., Brickman, P., & Bolen, D. (1975). Attribution versus Persuasion as a Means for Modifying Behavior. *Journal of Personality and Social Psychology*, 31(3), 430–441. doi:10.1037/h0076539.
- O'Connor, C. (2019). *The Origins of Unfairness: Social Categories and Cultural Evolution* (First edition). Oxford and New York: Oxford University Press.
- Peters, U. (2019). The Complementarity of Mindshaping and Mindreading. *Phenomenology and the Cognitive Sciences*, 18(3), 533–549. doi:10.1007/s11097-018-9584-9.
- Roberts, R. C. (1984). Will Power and the Virtues. *The Philosophical Review*, 93(2), 227–247. doi:10.2307/2184584.
- Snyder, M. (1984). When Belief Creates Reality. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 18, pp. 247–305). Cambridge: Academic Press.
- Snyder, M., & Klein, O. (2005). Construing and Constructing Others. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 6(1), 53–67. doi:10.1075/is.6.1.05sny.
- Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. Cambridge, MA: The MIT Press.
- Tanesini, A. (2020). The Gift of Testimony. *Episteme*, 1–18. doi:10.1017/epi.2019.52.

- Westra, E. (2020). Folk Personality Psychology: Mindreading and Mindshaping in Trait Attribution. *Synthese*, 198, 8213–8232. doi:10.1007/s11229-020-02566-7.
- Zagzebski, L. T. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.
- Zagzebski, L. T. (2012). *Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief*. Oxford and New York: Oxford University Press.
- Zawidzki, T. W. (2013). *Mindshaping: A New Framework for Understanding Human Social Cognition*. Cambridge, MA: MIT Press.
- Zawidzki, T. W. (2018). Mindshaping. In A. Newen, L. d. Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 735–754). Oxford: Oxford University Press.

T&F Proofs – Not for Distribution

5b Commentary from Ian James Kidd

Comments on Alessandra Tanesini, “Mindshaping and Intellectual Virtues”

IAN JAMES KIDD

Our epistemic characters and lives are complicated things. Looked at from one perspective, there is a story to tell about epistemic virtues, thinking and exploring, and the cooperative pursuit of truth, knowledge and understanding through systems of enquiry. Looked at from another perspective, it is a story of epistemic vices, obstruction and willed ignorance, and the individual and collective determination to turn away from epistemic goods. Each perspective is essential since each captures important aspects of our epistemic lives. Alessandra Tanesini does superb work in exploring that second perspective through her work on epistemic vices, a guiding theme of which is that we need a vice epistemology alongside a virtue epistemology and an anti-social epistemology to complement our social epistemology.

In her chapter, Alessandra argues that the epistemic virtues should be understood as epistemic character traits which have been “culturally selected for their social epistemological and practical benefits”. She elaborates this in reference to *mindshaping*, an account of folk psychology according to which we human beings are able to “typically able to predict others’ behaviour by correctly figuring out the independently formed mental states that guide that behaviour”. The ultimate aim of this activity is to *shape* the minds of others, by trying to get others to believe and desire in ways that are proper, and so on. This is true of attributions of epistemic character traits, including epistemic virtues, like open-mindedness, inquisitiveness and epistemic humility. For advocates of mindshaping, “whenever people attribute propositional attitudes, emotions, character or personality traits to human beings, what they are doing (irrespective of their intentions) has the primary function of shaping the minds of those to whom these features are attributed”.

A quibble is that this may seem unfalsifiable: no matter what you actually think you’re doing, you are engaging in mindshaping – you just don’t know it, or won’t accept or admit it. The quibble, though,

points to a more substantive idea I want to float. It is the idea that what really characterises many of our most meaningful interpersonal relationships is really mutual acts of *exploration*. Sometimes, we might be engaged in trying to *shape* the minds of others, since doing so is often appropriate or even obligatory. In other cases, though, something else is arguably fundamental that is (a) different from mindshaping, (b) incompatible with shaping the mind of the other, and also (c) inclusive of a variety of epistemic and moral virtues. I have in mind experiences of empathy.

Empathising with others is often understood in terms of simulationism, according to which empathy requires that two people have an experience in common. Empathising means modelling the mind of the other (see Goldman on mirroring and reconstruction “routes” into empathy). Simulationism, of this sort, is related to forms of folk psychology integral to forms of the mindreading account discussed by Tanesini. It has also been robustly criticised on several counts (for instance, replicating someone’s experiences isn’t the same as understanding them and modelling someone else’s first-person perspective may really be a way of imposing one’s own first-person perspective onto another). I want to focus, though, on a different conception of empathising, which privileges the *exploration* of others’ minds. I’ll sketch its details and then suggest that it offers a different way of thinking about the origin of epistemic virtues.

In the phenomenological tradition, empathy is understood, not as simulation, but as a distinctive kind of intentional state – as an experience of one’s own that presents someone else’s experiences *as* someone else’s. Empathising with someone isn’t a matter of simulating or modelling their experiences. It is a perception-like exploration of someone’s experiences, as disclosed in their embodied behaviours and interpersonal interactions – their moods, tone, demeanor, speech and so on. Specific acts of simulation may be an aspect of this process, of course, but only in a limited, secondary way alongside a diverse array of cognitive, affective, imaginative and moral achievements. Empathising with someone is better understood as the activity of exploring someone’s experience against the background of a shared social world – a common context of values, standards, commitments, a sense of salience and meaningful shared possibilities.

Empathising is only one of our interpersonal practices and therefore is not the whole of our interpersonal life. It is, though, a distinctive one that arguably represents a wonderfully human achievement – an accomplishment that shows our epistemic, moral, and imaginative capacities at their best. I want to suggest, though, that when empathising with others, what’s often more fundamental is empathetic understanding, something that must be achieved prior to any shaping of the person’s mind. Certainly, this seems the case when empathising with those experiencing grief, trauma, chronic illness and other painful life experiences, where

the crucial task is *appreciating differences* between one's own experience and that of the other while resisting the urge to assimilate the one to the other in a way that erodes the first-person distinctiveness of their experience. The therapist Carl Rogers describes this delicate feat:

To sense the client's anger, fear, or confusion as if it were your own, yet without your own anger, fear, or confusion getting bound up in it, is the condition we are endeavoring to describe. When the client's world is this clear to the therapist, and he moves about in it freely, then he can both communicate his understanding of what is clearly known to the client and can also voice meanings in the client's experience of which the client is scarcely aware.

(Rogers 1957, 99, my emphasis)

In these cases, the immediate task is to explore the shape, contours, rhythms, and character of the other person's experience – to enter into their world, as it were, and with that person come to explore that world in all its particularity and difference. At this early point, *shaping* would seem premature, especially if the person is experiencing the disruption and uncertainty integral to so many painful life experiences.

This creates special roles for a variety of virtues, including attentiveness to differences, cautiousness, self-restraint, humility, openness and patience. To explore the very different world of another person, to resist the urge to impose meaning and structure onto it from the outside, to restrain a desire to assimilate their experience to one's own, to maintain a disciplined commitment to a style of interpersonal epistemic engagement that is more perambulatory than probing ... all of this and more requires a whole array of very specifically inflected epistemic and moral virtues. If successfully exercised, such virtues enable a richly empathetic understanding of another person's distinctive world of experience that is well characterised by Knud Løgstrup:

By our very attitude to one another we help to shape one another's world. By our attitude to the other person we help to determine the scope and hue of his or her world; we make it large or small, bright or drab, rich or dull, threatening or secure. We help to shape his or her world not by theories and views but by our very attitude toward him or her. Here lies the unarticulated and one might say anonymous demand that we take care of the life which trust has placed in our hands.

(Løgstrup 1997, 18)

If these thoughts are on the right track, then mindshaping may not be the best way to think about at least one important aspect of interpersonal life – the empathetic project of trying to enter into and explore

the distinctiveness of and differences among the worlds of experience inhabited by so many of our fellows suffering some of the worst things a human being can endure.

References

- Løgstrup, Knud (1997) *The Ethical Demand* (Notre Dame: University of Notre Dame Press).
- Rogers, Carl (1957) 'The Necessary and Sufficient Conditions of Therapeutic Personality Change', *Journal of Consulting Psychology* 21: 95–103.

T&F Proofs – Not for Distribution

5c Commentary from Thi Nguyen

Comment on Tanesini's "Mindshaping and Intellectual Virtues"

THI NGUYEN

According to Tanesini, the formation of virtues and vices – of character – should be treated as a kind of mindshaping. I worry, however, that her particular take on mindshaping encourages us to export some problematic presumptions into our theory of the intellectual virtues.

One of the primary goals of mindshaping, says Tanesini, is social convergence:

Mindshaping activities are uniquely well-suited to facilitate the kind of maximal mutual intelligibility instrumental to solving coordination problems. It is much easier to coordinate one's actions with a person who conforms to what we expect of them than with one who does not. The effect of mindshaping is to bring oneself and others to behave in accordance with the same norms and to be intrinsically motivated to follow them.

Let me emphasise the key idea: We solve coordination problems through maximising mutual intelligibility, achieved by *convergence on the same norms*. Tanesini then suggests that education and character formation be treated as a kind of mindshaping. If we export this convergence-centric conceptualising of mindshaping to a virtue approach, then we should expect character mindshaping to also aim at convergence – at the creation of the *same types* of character in everybody, characters which follow the same norms and are mutually intelligible to one another.

But, in my mind, a virtue-based account is desirable in part because it can depart from this universalising, legalistic framework. Legalistic approaches create coordination through enforced convergence on the same norms. And those ethical systems that are founded in a legalistic conception of morality idealise the same kind of normative convergence.

But virtue theory is, to my mind, so compelling precisely because it is open to a more pluralistic vision of communal moral life. It permits

imagining a society that is coordinated through a balance of profoundly different moral characters. One person might be the fierce and enraged warrior for social justice; another might be the gentle and empathetic listener; another a nitpick-y conceptual analyst of ethical concepts. They all have something to contribute. (At least, the first two certainly do.) And to contribute, their actions need not be wholly intelligible to one another. In other words, virtue theory is primed to support a rich moral community, achieved through the division of moral labor.¹ And intellectual virtue theory seems particularly exciting to me because, for similar reasons, it seems richly compatible with various views that epistemic communities function better when there is a vast diversity of intellectual characters, interacting (Kitcher 1990).

Of course, we could also imagine a mindshaping story where we mindshaped in pursuit of a rich and balanced mix of diverse intellectual characters. But I urge a bit of caution here about emphasising the *convergence* aspect of mindshaping. There are more ways to coordinate than convergence.

Note

1 For an opening discussion of the division of moral labor, see Nguyen (2021).

References

- Kitcher, Philip. 1990. "The Division of Cognitive Labor." *The Journal of Philosophy* 87 (1): 5–22.
- Nguyen, C. Thi. 2021. "Transparency Is Surveillance." *Philosophy and Phenomenological Review*. <https://doi.org/10.1111/phpr.12823>

5d Alessandra Tanesini's Response to Commentaries

A Rejoinder to Nguyen and Kidd

ALESSANDRA TANESINI

In their insightful critical commentaries on my chapter Thi C. Nguyen and Ian James Kidd seek to emphasise the role of moral and intellectual virtues in fostering and appreciating human diversity in cognitive proclivities and character traits. They note that my chapter places too much emphasis on homogeneity, moulding and imitation, and seek to balance it. Kidd remarks that empathy is about openness to exploring another's mind; Nguyen points out that cooperation does not require homogeneity in dispositions among all actors. Let me begin this rejoinder by way of a partial concession. It is true that in my chapter I have somewhat over-emphasised homogeneity and convergence at the expense of diversity in the service of distribution of cognitive labour. I have also highlighted the ways in which mindshaping might resemble indoctrination. I have foregrounded these aspects for a reason that I would like to make explicit in this response.

In the chapter, I argue that educative practices that foster the formation of character traits, and especially of those that are identified as virtues, make the communities that have developed them more successful at solving cooperation problems. Nguyen is right that many of these problems are solved by means of dividing the community into groups each of which specialises in a specific activity. This is a point I also make towards the end of my chapter. It is true that the view that the cultural evolution of virtues is in the service of cooperation does not require absolute convergence over one specific set of dispositions. On the contrary, one would expect the proliferation of diverse characters and skills each suited to diverse roles. Nevertheless, cooperation requires mutual intelligibility and to this extent, a certain amount of cognitive similarity is required to foster mutual understanding. I do not take my disagreement with Nguyen in this regard to be substantive.

The chapter's focus on character formation as mindshaping practices that rely on imitation to create somewhat homogenous cognitive architectures is a provocation intended to highlight the continuities between

character and virtue education and what Foucault (1979) has described as exercises of disciplinary power. In *Discipline and Punish* Foucault examines institutions, such as schools and prisons, and practices that shape the minds and bodies of children and citizens turning them into “docile bodies” well suited to function in capitalist industrial societies. The description of individuals as docile, of course, is intended to highlight how disciplinary practices erode the autonomy of those they impact by changing their personalities and behaviours in ways that are not necessarily conducive to their flourishing. But, as Foucault also emphasised, these practices by creating new kinds of personality also enable new kinds of autonomous agency. That is, disciplinary power serves both to limit some freedoms, and to create others. Disciplinary practices are not intrinsically bad because they are not always in the service of social injustice. The same, as McGeer (2019) has remarked, applies to mindshaping, and I wish to add, to virtue education and virtue attribution.

By showing that virtue education is an exercise in mindshaping that encourages the development of dispositions of self-discipline and self-control I intend to highlight the dangers inherent in this pedagogical practice. These are dangers that are orthogonal to the reasons to promote diversity and the worries about homogeneity justly highlighted by Nguyen. These dangers are exemplified by past discussions of feminine virtues of humility and masculine warrior virtues of courage or integrity that contributed to the re-enforcement of unfair distributions of cognitive, emotional, and material labour by gender. The history of virtue theory and virtue talk is chequered, since they have often been deployed to trench inequity. My chapter is intended to explain the social epistemic value of shaping human minds in the direction of intellectual virtues in a manner that also highlights how virtue attribution can be put to work in the service of indoctrination and of other unjust practices. It is also intended to be alert to the possibility highlighted by O’Connor (2019) that social epistemic success in some instances might be gained at the expense of justice.

This is a genuine risk because sometimes solutions to problems that make societies more successful as a whole are achieved at the cost of unfair distributions of burdens among its members (O’Connor, 2019). This is not a mere theoretical possibility but it is frequently an actuality that has disadvantaged subordinated social groups. Philosophical accounts of which virtues are best suited for individuals from diverse walks of life have in the past been instrumental in bringing about such unfairnesses. The account of virtues as the product of mindshaping practices offered in my chapter is intended to provide the theoretical background against which the epistemic and moral advantages and dangers inherent in character formation can appear in stark relief.

For this reason, and also because in my chapter I explicitly exclude any commitment to the radical view that mindshaping practices predate

mindreading abilities, I do not take Kidd's careful examination of the virtues of empathetic understanding to stand in opposition to my view. It is perfectly possible that some ability to mindread empathetically is required for mindshaping to be effective; but, it is equally possible that mindshaping practices are instrumental in the development of the kind of cognitive abilities and dispositions involved in empathetic understanding. Be that as it may, Kidd, like Nguyen, focus on the value of virtue cultivation in the service of knowledge and understanding. I share their point of view, but its appreciation should not lead us to forget that character education, especially when carried out in societies marked by widespread inequity, is easily co-opted in the service of deepening injustice rather than relieving it. This ease of co-option, my chapter indicates, is a by-product of the deeply collective epistemic function of intellectual virtues.

References

- Foucault, M. (1979). *Discipline and Punish: The Birth of the Prison* (A. Sheridan, Trans.). New York: Vintage Books.
- McGeer, V. (2019). Mindshaping Is Inescapable, Social Injustice Is Not: Reflections on Haslanger's Critical Social Theory. *Australasian Philosophical Review*, 3(1), 48–59. doi:10.1080/24740500.2019.1705231.
- O'Connor, C. (2019). *The Origins of Unfairness: Social Categories and Cultural Evolution* (First edition). Oxford and New York: Oxford University Press.

T&F Proofs – Not for Distribution

Part II

**Individual Virtues
and Vices**

T&F Proofs – Not for Distribution

T&F Proofs – Not for Distribution

6 The Vices and Virtues of Extremism

Quassim Cassam

1

When a person is labelled as an ‘extremist’ it is natural to suppose that this act of labelling serves at least two purposes: to describe and to evaluate.¹ The implied evaluation is usually negative but what is the label’s descriptive content?² Does it even have a definite descriptive content and is there a real feature of some individuals that answers to this content? I will argue that one real feature of individuals that the ‘extremist’ label picks out is their *mindset*. The idea that there is an extremist mindset is not new but existing accounts of this mindset are sketchy. This is partly a reflection of the fact that the idea of a mindset is far from clear. Some accounts of the extremist mindset represent it as a belief or way of believing. Others describe it as a way of thinking or thinking pattern. There is also the idea that it is an attitude or attitude disposition. One challenge, therefore, is to clarify the general idea of a mindset and, specifically, the notion of an extremist mindset.³

People are not the only entities that are described as extremists. As well as beliefs, ways of thinking and attitudes, this label also applied to ideologies, behaviour, policies, groups and movements. On a suitably expansive conception of ideology, ideologies are mindsets and extremist ideologies are, or give expression to, an extremist mindset.⁴ Extremist movements or groups can be understood as ones that subscribe to and are motivated by extremist ideologies. It follows that the extremist mindset also underpins extremism at the level of movements or groups. Extremist policies can also be understood as an expression of an extremist mindset. It should be noted that extremism is often characterised in terms of a commitment to violence. On the account given here, violent extremism is one form of extremism but extremism needn’t be violent.

Is an extremist mindset necessarily bad? Are there circumstances in which a person or group might deserve to be commended rather than condemned for being extremist? On the one hand, there is a strong intuitive case for viewing the extremist mindset as epistemically, morally and politically vicious.⁵ On this view, extremism is to be countered by encouraging the development of a range of anti-extremist virtues. On

the other hand, it might also be held that extremism can be positive when it is extremism in support of a just cause. It has been suggested, for example, that the suffragettes were extremists but is this not a case in which extremism was justified? Extremism is partly a matter of being unwilling to compromise and there are surely some issues in relation to which there is no room for compromise. Votes for women is one such issue. It remains to be seen, however, whether such examples of the supposed virtues of extremism are compelling.

One welcome consequence of a mindset approach to extremism is that the classification of a person or group or ideology as extremist isn't simply a matter of opinion or an exercise in political rhetoric. No doubt the label 'extremist' is often applied for narrowly political reasons but if it is an objective matter whether someone has an extremist mindset then it is also an objective matter whether that person is an extremist. This is one sense in which the label 'extremist' picks out something real. Mindsets aren't fictions. This is not to deny, however, that having an extremist mindset is a matter of degree. A person or group can be more or less extremist. Extremism isn't all or nothing, and one evaluative question is whether extreme extremism is significantly worse than what might be called, somewhat oxymoronically, more moderate forms of extremism.

A test for any account of extremism is whether it delivers the correct verdicts about specific individuals or organisations. For example, an account of extremism is unacceptable if it implies that organisations like ISIS (the Islamic State of Iraq and Sham) or individuals like the Norwegian mass murderer Anders Breivik are not extremists.⁶ In fact, there is no danger of the mindset approach delivering such perverse verdicts. The mindset of ISIS and its leaders is a paradigm case of an extremist mindset. Since extremism can be non-violent, there is also scope for examining the role of the extremist mindset in non-violent political conflicts. For example, it is arguable that non-violent extremism has played a role in the Brexit debate in the United Kingdom. To the extent that mindsets are psychologically real, a further question for the mindset approach is whether it accords with the empirical psychological evidence. As it happens, there is psychological evidence of a 'Militant Extremist Mindset' (MEM) and the papers in which this evidence is reported cast further light on the concept of a mindset.⁷

The discussion below will proceed as follows: Part 2 will explain the idea of a mindset and develop the notion of an extremist mindset. As understood here, the extremist mindset is constituted by, among other things, a distinctive set of attitudes, pre-occupations, emotions, and thinking patterns. These attitudes, pre-occupations, emotions and ways of thinking are liable to cause types of behaviour that are associated with, though not uniquely, extremism. The mindset approach to extremism will be compared to other approaches and be shown to be consistent

with the psychological evidence about extremism. It will also be shown to deliver the correct intuitive verdicts in particular cases.

Part 3 will explore the sense or senses in which extremism is epistemically, politically and morally vicious. This will necessitate a brief discussion of what counts as an epistemic, political or moral vice. This will also be the place for a discussion of the supposed upside of extremism in relation to just causes. Regardless of whether the suffragettes were extremists, is it not conceivable that extremism might be politically virtuous, that is, better able to advance just causes than moderation? Conceivable, perhaps, but in practice the political harms done by extremism far outweigh any supposed benefits. The determination, implacability and tenacity displayed by campaigners for just causes should not be confused with extremism.

Part 4 will explore the causes and sources of extremism as well as potential antidotes. Is the extremist mindset a personality trait or an acquired or inculcated politico-psychological posture? If it is acquired then it will need to be explained how it is acquired. One notion that is sometimes employed to explain the process of becoming an extremist is that of *radicalisation*. The suggestion is that people become extremists either by self-radicalising or being radicalised by others. Following a brief discussion of this suggestion, I will conclude by identifying some of the anti-extremist virtues that might have a part to play in countering extremism. If there are such virtues, then the practical challenge is to identify ways of educating for them.

2

The concept of a mindset will be familiar to some readers from the work of Carol Dweck. Mindsets in Dweck's sense are 'just beliefs' (2012, 16). So, for example, what Dweck calls the 'growth' mindset is 'based on the belief that your basic qualities are things you can cultivate through your own efforts' (2012, 7). As understood here, mindsets are not just beliefs, and there is no 'extremist mindset' if that means that there is a single belief that all extremists have. Mindsets are closer to world views or frameworks through which the world is viewed and understood. They shape our beliefs and filter our perception of reality. In this respect, there is a parallel with Kant's categories, but mindsets aren't just concepts, any more than they are just beliefs.

Mindsets are partly constituted by pre-occupations. One's beliefs are relevant to one's mindset to the extent that they underpin and explain one's pre-occupations. Two key extremist pre-occupations are *persecution* and *purity*.⁸ Extremists are typically pre-occupied with the idea that they belong to a persecuted or victimised group, and convince themselves that extreme measures are called for in response. Nazi propaganda made much of the threat to Germany posed by a supposed Jewish

world conspiracy, and there are many other examples of extremists with lurid fantasies of victimisation and persecution. Anders Breivik justified the killing of 77 people in Norway in 2011 partly on the grounds that Christian civilisation was threatened by Islam. The threat of subordination to Islam is also a part of the mindset of Buddhist extremists in Myanmar, and many Muslim extremists see Islam as threatened by the 'Crusader' West.

These examples might be thought to imply that the pre-occupation with persecution that plays a significant role in the extremist mindset is baseless, hence the characterisation of this pre-occupation as relying on lurid *fantasies* of persecution and victimisation. But what if the persecution is real? Would this then invalidate the description of mindsets that are pre-occupied with persecution as 'extremist'? Not necessarily, since other elements of an extremist mindset might be present in a given case even if the persecution is genuine. It is still plausible, however, that the extremist mindset is paradigmatically pre-occupied with non-actual persecution. Where there is genuine persecution, such as the persecution of the black population of South Africa under apartheid, it might be appropriate to refrain from describing those engaged in a struggle against such oppression as extremists, though much will depend on their other pre-occupations and other aspects of their mindset.

The purity with which extremists are pre-occupied can take many different forms: racial or ethnic, religious, ideological, and so on. For the Khmer Rouge in Cambodia, the pursuit of ideological purity was bound up with 'a racialist project of ethnic purification' (Kiernan 2008, xxx). For ISIS, what matters is religious purity. It sees itself as defending and promoting a pure and unadulterated form of Islam, grounded in a literal reading of the Koran. Carolin Emcke highlights ISIS's 'cult of purity' (2019, 102) and its perception of itself as the only 'authentic' Muslims. Their lack of purity justifies the targeting of the polluted and impure by all available means, including violence and intimidation. For all the ideological differences between the Khmer Rouge and ISIS, their pre-occupation with purity points to a shared extremist mindset. Given the extent to which extremists are pre-occupied with purity, it comes as no surprise to find many of them engaged in acts of ethnic, ideological or religious 'cleansing'.⁹

The attitudinal components of the extremist mindset are easily identified. One's attitude towards something is one's stance or posture towards it. A key extremist attitude, and one that flows from its pre-occupation with purity, is its attitude towards compromise. Extremists are bitterly opposed to compromising their ideals and objectives.¹⁰ As they see it, compromise is incompatible with purity, and this explains their perception of compromise as a form of betrayal that can never be countenanced. As well as flowing from its obsession with purity, extremism's uncompromising attitude is related to its Manichaeism. If the

world is divided into good and evil, believer and infidel, and one thinks of one's opponents as utterly depraved and misguided, then negotiating or compromising with them would amount to negotiating or compromising with evil.

Extremism's view of compromise is a reflection of its certainty in its own rectitude and the complete absence from its mindset of any element of self-doubt. Certainty and absence of self-doubt are epistemic postures, attitudes towards one's own epistemic standing and that of one's principles and commitments. The extremist's certainty is subjective, though taken to be objective. The extremist is totally convinced of the correctness of his principles even though, objectively speaking, there is plenty of room for doubt. Certainty is not necessarily a sign of extremism. Being certain that two plus two is four or that slavery is indefensible does not make one an extremist. The extremist's psychological certainty pertains to matters in regard to which an absence of doubt is inappropriate. The extremist is not only doubt-free in relation to his doctrinal commitments but also in relation to his own grasp of the truth. It is not just doubt that he lacks, but *self*-doubt. Like the ISIS supporters described by Graeme Wood in his study of the Islamic State, he revels in his self-confidence and luxuriates in the 'banishment of uncertainty' (2018, 103).

Another characteristically extremist attitude is a kind of indifference or insouciance about the practical implications or consequences of their policies. This practical indifference is helpfully characterised in the following terms by Scruton: extremism takes a political idea to its limits, 'regardless of unfortunate repercussions, impracticalities, arguments, and feelings to the contrary, and with the intention not only to confront, but to eliminate opposition' (2007, 237). Extremists are not deterred by the notion that their approach will have catastrophic consequences for large numbers of people. For example, the Khmer Rouge was indifferent to the fact that their policies would result in the death by starvation of millions of Cambodians. For the extremist, such consequences are a price worth paying for ideological purity. Indeed, the true extremist goes even further than the character described by Scruton; the Khmer Rouge didn't even regard the repercussions of their murderous policies as unfortunate. The extremist's motto is: you can't make an omelette without breaking eggs.

The practical indifference that is an element of extremism is the essence of fanaticism. Fanatics have been described as 'aggressive and potentially violent ideologues' (Saucier et al. 2009, 268). An ideologue is supposedly someone with 'a high degree of commitment to an ideology' (ibid.). It remains to be seen how talk of degrees of commitment is to be cashed out. Meanwhile, a natural thought is that the higher one's degree of commitment to a principle the less one's concern about any unfortunate repercussions or impracticalities. The fanatic sees any unfortunate repercussions as a price worth paying. A person who is not practically

indifferent is not a fanatic even if they display several other characteristics of the extremist. In practice, however, extremism and fanaticism go hand in hand.

The extremist's unwarranted psychological certainty is usually sustained by high levels of closed-mindedness and dogmatism.¹¹ These can either be conceived of as character traits or as attitudes. For present purposes they are attitudes. Closed-mindedness consists in, among other things, having a poor appreciation of perspectives that are different from one's own, a high degree of intolerance of alternative perspectives, and a tendency to reject information that is inconsistent with what one already believes. Dogmatism pertains to one's specific doctrinal commitments rather than one's epistemic conduct generally. It is an irrational commitment to a fundamental doctrine.¹² It stands to reason that extremists who are supremely convinced of the correctness of their doctrines will be hostile to alternative points of view. To the extent that these doctrines are themselves baseless, the extremist's commitment to them is also likely to be irrational. The question of what, in general, makes a commitment irrational cannot be considered here.

The emotional components of the extremist mindset include hatred, fear and self-pity.¹³ Hatred of the ideological or religious Other is a major driving force of extremism. Extremists don't just see individuals with a different take on reality as people with whom they disagree. As noted above, they see them as evil and depraved. Extremism's hatred is tied to its sense of certainty. As Emcke notes, 'hating requires absolute certainty' because 'you cannot hate and be unsure about hating at the same time' (2019, xi). In its most extreme form, extremist hatred results in the 'othering' of one's opponents. Othering is 'the attribution of relative inferiority and/or radical alienness to some other/ out-group' (Brons 2015, 83). The 'other' is regarded as barely human, as an entity that can be 'disregarded or denounced, injured or killed, without fear of punishment' (Emcke 2019, xii). This is ISIS's attitude towards Jews, Christians and Shia Muslims, and it is how Buddhist extremists in Myanmar see the Rohingya.

Extremism's hatred of the other is typically grounded, at least in part, in fear.¹⁴ Fear of the other is related to extremism's pre-occupation with persecution by the other. One curious feature of this pre-occupation is that the persecution is usually imaginary. In most instances, the feared other poses no real threat to the extremist and is, indeed, itself a victim of persecution by extremists. Nevertheless, extremists like to think of themselves as victims. What Ruth Ben-Ghiat describes as the 'cult of victimisation' is at the core of their identity and explains the key role of self-pity in the extremist mindset.¹⁵ As O'Toole notes, self-pity combines 'a deep sense of grievance and a high sense of superiority' (2019, 3). In Myanmar, Buddhist extremists have a deep sense of grievance against the Rohingya, but take the inferiority of the Rohingya for granted.

Having identified its pre-occupations, attitudes and emotions it remains to identify the styles of thinking or thinking patterns that are associated with an extremist mindset. Among these thinking patterns are catastrophic thinking, utopian thinking, apocalyptic thinking and conspiracy thinking. The blandest form of catastrophic thinking is the tendency to exaggerate the negative consequences of our life situations.¹⁶ Extremist catastrophising goes well beyond that; it usually involves the idea of an impending catastrophe for the extremist's in-group that can only be averted by extreme measures.¹⁷ The promise of extremism is that it is the route to utopia or, in the case of some religious extremists, paradise. Apocalyptic thinking consists of the tendency to think of the ideal end-state as attainable only after an apocalyptic battle with the forces of evil. This form of apocalyptic thinking is, for example, integral to the mindset of ISIS, which has even identified a town in Syria as the venue for one of its final battles.¹⁸

The relationship between extremism and conspiracy thinking deserves more attention than it is possible to give it here. For present purposes, it is sufficient to note that, as a matter of historical fact, conspiracy theories have often been used to promote extremism.¹⁹ Right-wing and left-wing extremists have both relied on the myth of a world Jewish conspiracy to justify their anti-Semitism, and both Hitler and Stalin were conspiracy theorists. Just as conspiracy thinking promotes extremism, extremism makes one more liable to engage in this type of thinking. There are extremists who are not a conspiracy theorist but the point of talking about an extremist mindset is not to identify strictly necessary conditions for extremism. The attitudes, pre-occupation, emotions and thinking patterns identified here are not all required for a mindset to qualify as extremist, but a reasonable number must be present. It is in this sense that these things are constitutive of extremism or the extremist mindset.

How has the extremist mindset been identified? On what basis is a given attitude or pre-occupation or emotion or thinking pattern said to be part and parcel of this mindset? The nature of the extremist mindset cannot be identified without reference to the mindset of actual extremists, that is, the mindset of individuals or groups that are widely regarded as extremist. This is the methodology employed in recent empirical work on the Militant Extremist Mindset (MEM). Specifically, it has been suggested that the description of this mindset should be 'grounded on "themes" (recurrent patterns of thinking, feeling and behaving) based on explicit statements found in primary sources and characterising at least three different extremist groups' (Stankov, Saucier and Knežević 2010, 71). More recent work on the MEM has identified a total of 16 key themes, including several that I have identified as components of the extremist mindset.²⁰ These themes have been identified by induction rather than by *a priori* conceptual analysis.

A key ingredient of MEM that has not been featured in the account that I have given is pro-violence, the belief that violence is a useful and legitimate means of achieving one's goals. The omission of pro-violence and, indeed, actual violence from the extremist mindset is a reflection of the distinction between extremism and militant extremism.²¹ Extremism need not be violent or even pro-violence even if, in practice, a great deal of extremism is both of these things. The othering of out-groups can and often does result in violence but there are many non-violent means of oppressing the Other. It should be conceded, however, that the extremist individuals, groups and organisations on which I have based my account – anti-Rohingya extremists in Myanmar, the Khmer Rouge, Anders Breivik and ISIS - are all violent. To the extent that such individuals and organisations are the basis of one's understanding of extremism, there is no danger of the resulting account of extremism delivering the perverse verdict that they are not extremist.

The mindset approach to extremism contrasts with several other approaches. On what might be called a ideological conception of extremism, an extremist position 'falls somewhere near the end or fringe of something close to a normal distribution' (Nozick 1997, 296) along some salient political dimension. Left-Right is one such dimension but there are others, and positions that were once viewed as extreme 'later often come to be viewed as somewhere in the center' (Nozick 1997, 296). On this conception, an extremist *move* can be defined as 'a move away from the centre and towards the extreme in some dimension' (Wintrobe 2010, 25). On a *modal* conception of extremism, in contrast, what counts is not *what* one believes but *how* one believes. Extremism is essentially 'a characteristic of the way beliefs are held rather than their location along some dimension' (Breton, Galeotti, Salmon and Wintrobe 2010, xiii). A *methods* conception of extremism holds that being an extremist is a matter of being willing to use or endorse extreme methods. Such methods are usually understood as violent, and it is in the methods sense that many terrorists are extremists.

Of these three conceptions, the second is the closest to the mindset approach. A question about modal extremism concerns its understanding of 'the way beliefs are held'. This can be understood as a reference to the strength or intensity of the extremist's beliefs. The most intense beliefs, on this view, are accompanied by the strongest or most intense feelings of conviction. Yet, as Ramsey notes, 'the beliefs which we hold most strongly are often accompanied by practically no feeling at all; no one feels strongly about what he takes for granted' (1931, 169). On an alternative reading, the strength of one's beliefs is a function of one's willingness to give them up or compromise them. There are many reasons why a particular belief might be treated as immune to revision. Beliefs that help to define one's world view or sense of identity are not easily given up. The problem with extremists is not that they have bedrock or 'hinge'

beliefs in this sense; we all do.²² The problem is that the specific principles or propositions they take for granted are in fact highly contentious and far from unproblematic.

Holding onto one's beliefs and principles in a rigid and uncompromising manner is one element of the extremist mindset but there is much more to it than that. Given that having an extremist mindset is a matter of having certain specific pre-occupations, it is not possible to understand extremism in purely formal terms, in terms of how one believes rather than what one believes. Extremism is, to some extent, a matter of what one believes. For example, a pre-occupation with loss of purity is a substantial rather than a purely formal feature of extremism. Beliefs about purity and victimhood are bedrock extremist beliefs. They, together with the other features of the extremist mindset, indicate that extremism is an ideology in its own right, and not just a way for one to hold onto one's political or other beliefs regardless of their content.

This has a bearing on the question of whether extremism is compatible with any political philosophy. For example, is it possible for one to be a liberal extremist? If extremism is simply a matter of 'the way beliefs are held' then there is no reason in principle not to classify some liberals as extremists. After all, liberals can be just as uncompromising about their core beliefs as those who are more usually classified as extremists. Yet it would be perverse to characterise uncompromising liberals as extremists if they are not pre-occupied with victimhood or purity and do not have an extremist thinking style. When a person is described as an extremist it is usual to ask 'an extremist what?'. This is a legitimate question to ask, insofar as extremism can take many different forms. However, it does not follow that extremism can be combined with *any* political or religious beliefs, or that describing a person as an extremist on its own implies nothing about their substantive commitments. There must be some such commitments, or least pre-occupations, if this label is to apply. The complete absence of hatred and a lack of practical indifference are also incompatible with extremism. Liberals who do not hate their opponents, do not engage in othering, and are not practically indifferent are just not extremists, regardless of how rigidly they hold on to their core values. This is the truth in the ideological conception of extremism: people whose politics place them close to the centre of a normal distribution are highly unlikely to have the substantive ideological commitments required for them to qualify as extremists.

3

What is wrong with having an extremist mindset? Is such a mindset morally, politically or epistemically vicious? It is easy to make the case that extremism is a moral failing. It is a moral failing to be indifferent to the consequences of one's actions and policies for other human beings.

It is a moral failing to engage in the ‘othering’ of people with whom one disagrees, and it is morally indefensible to be motivated by a concern for ideological, religious or racial purity. Whatever else there is to say about virtues, they are ‘in general beneficial characteristics, and indeed ones that a human being needs to have, for his own sake, and that of his fellows’ (Foot 1978, 3). An extremist mindset is not, in general, beneficial, even if there are circumstances in which it might be. Extremism is not a virtue, and the harms done by extremists over the years suggest that it is, in fact, a vice.

For present purposes, epistemic vices can be understood as character traits, attitudes or thinking styles that get in the way of the gaining, keeping or sharing of knowledge.²³ As well as getting in the way of knowledge, epistemic vices are personal qualities that merit criticism or blame. The closed-mindedness and dogmatism that characterise the extremist mindset both get in the way of knowledge and merit criticism. The various thinking patterns that are part and parcel of the extremist mindset are no less epistemically problematic. Conspiracy thinking, or what psychologists refer to as a ‘conspiracy mentality’, leads extremists to endorse fallacious or even contradictory conspiracy theories.²⁴ Catastrophic thinking is an obstacle to knowledge of one’s actual situation or prospects, and the apocalyptic thinking which groups like ISIS find so irresistible further weakens their grip on reality. It is also arguable that one is responsible for one’s own thinking and attitudes.²⁵ In that case, there is no prospect of extremists being immune to blame or criticism for their extremist thinking patterns and attitudes on the basis that they aren’t responsible for them.

Political vices have been defined as ‘persistent dispositions of character and conduct that imperil both the functioning of democratic political institutions and the trust that a diverse citizenry has in the ability of those institutions to secure a just political order of equal moral standing, reciprocal freedom, and human dignity’ (Button 2016, 1). One might quibble about some aspects of this definition. One might want to allow attitudes, thinking patterns and even emotions to count as political vices. There is also the question of whether political vices should be identified exclusively by reference to their effect on *democratic* political institutions or, for that matter, by reference to their effect on a nation’s political *institutions* rather than its political culture more generally. The basic point, however, is that political vices are *politically* damaging. One of the effects of extremism, indeed one of its intended effects, is polarisation. If extremism causes polarisation and the latter is politically damaging then that is one reason to classify extremism as politically vicious. No doubt there are plenty of others.

The claims that extremism causes polarisation and that the latter is politically damaging will not be defended in any detail here, though both seem obvious enough. The recent history of the United States and

United Kingdom amply demonstrates that polarisation is politically dysfunctional and causes severe difficulties for political institutions that were designed on the assumption of a broad consensus about fundamental values. The role of extremism in causing polarisation follows from its pre-occupation with purity, its tendency to engage in othering and its propensity for conspiracy thinking. In a recent analysis of extremism, J. M. Berger defines it as ‘the belief that an in-group’s success or survival can never be separated from the need for hostile action against an out-group’ (2018, 170). If this belief is part of the extremist mindset, then possession of that mindset is almost bound to cause the in-group and out-group to polarise.

Yet this line of argument faces the following apparently seductive response: there are surely circumstances, including some that are far from unusual, in which extremism is the only way to achieve worthy and democratically desirable objectives. In such cases, it can be an asset to have an extremist mindset, and there are plenty of examples of extremists who have done more good, politically speaking, than their more moderate allies. Indeed, not only is it possible that extremists are more effective than moderates but also that political actors at the far end of the extremist spectrum are even more effective than more ‘moderate’ extremists. If extremism can be politically beneficial in circumstances that are far from unusual, does this not call into question the idea that extremism is politically vicious?

Consider, again, the case of the suffragettes. On one view, they (or some of them) were extremists who campaigned successfully for votes for women, and their extremism was a significant factor in explaining their success. Since their extremism was politically effective and in a just cause, there is no reason to regard it as politically vicious. As it helped to overturn an obvious injustice – discrimination against women – it can also be regarded as morally virtuous rather than vicious. Finally, their extremism was epistemically beneficial to the extent that it gave them a clear insight into social and political injustices that were invisible to more moderate political opinion. In a similar vein, it might be said that the extremism of the African National Congress (ANC) in its battle against apartheid was justified and necessary; there was little hope of overthrowing apartheid by moderation. The ANC, with its extremist mindset, saw what needed to be done and did it. This, therefore, *looks* like another case in which extremism proved morally, politically and epistemologically beneficial.

In what sense were the suffragettes and the apartheid era ANC ‘extremists’? The usual explanation refers to the means or methods they employed. The ANC was engaged in armed struggle against the South African government and carried out acts of terrorism for which it was later held to account by its post-apartheid Truth and Reconciliation Commission. Some of the ANC’s terrorist acts resulted in the deaths of

civilians, including children. The bombings and arson carried out by the suffragettes were also clearly terrorist acts, regardless of whether they were justified.²⁶ It follows that both the ANC and the suffragettes were extremists in the methods sense. What is less clear is whether their use of extremist methods was effective. It is arguable that terrorist acts carried out by the ANC contributed little to the overthrowing of apartheid. The case of the suffragettes is even more complicated because, aside from questions about the effectiveness of their methods, there are also questions about their cause: unlike the ANC, they were not campaigning for universal adult suffrage.²⁷

The present question is not whether *terrorism* is politically, morally or epistemically vicious but whether an extremist mindset is vicious in any of these senses. Just as it is possible to have an extremist mindset without condoning or using violence, so it is possible to be pro-violence without having an extremist mindset. There is little evidence, for example, that Nelson Mandela or other senior members of the ANC had an extremist mindset despite being pro-violence, in the sense that they argued for an armed struggle against apartheid.²⁸ They did not engage in othering, were not pre-occupied with purity, and were responding to actual as distinct from imaginary oppression. People with an extremist mindset would not have set up a Truth and Reconciliation Commission after victory. If this is right, then the ANC provides no support for the idea that an extremist mindset can be beneficial or virtuous. Its leadership did not have an extremist mindset, and its greatest achievements would not have been possible with such a mindset. Similarly, it is hard to make the case that the suffragettes' extremist mindset – if they did indeed have such a mindset – contributed to their achievements.

The point of these considerations is not to suggest that it is absolutely inconceivable for an extremist mindset to be beneficial. The point is rather to suggest that convincing examples of this are much harder to find than one might suppose. Whether or not an extremist mindset is *invariably* harmful, the moral, political and epistemic harms that it *normally* causes are both systematic and predictable. This is enough to justify the classification of it as a moral, political and epistemic vice. The contrary view is sometimes based on a simple misreading of examples, such as those discussed above, of supposedly benevolent extremism, and sometimes on another simple misunderstanding: it is true that successful political campaigns against injustice require such qualities as determination, implacability and tenacity, and that many extremists have these qualities. However, it is possible to have these qualities without having an extremist mindset and the benefits of determination, implacability and tenacity in a just cause are more than likely to be cancelled out by the vicious aspects of such a mindset. There is therefore no reason to revise the initial verdict that extremism is a vice.

The remaining question is: how does a person come to have an extremist mindset? Is it an innate personality trait or is it acquired?²⁹ If it is acquired, how is it acquired, and what can be done to counter it? The empirical work in this area tends to focus on MEM. One view is that in the right conditions anyone is capable of becoming a militant extremist because MEM draws on ‘certain natural human tendencies’ (Saucier et al. 2009, 257). On the other hand, in a given context some individuals ‘may be more prone than others to take on this mind-set’ (Saucier et al. 2009, 257). How are such variations to be accounted for? Psychopathy might be one factor. This is taken to be a trait consisting of four characteristics: callousness, manipulativeness, lack of inhibition and anti-social behaviour. Sadism is another potentially relevant factor. According to one study, ‘proviolence was found to be predicted by sadism and psychopathy’ (Međedovic and Knežević 2018, 99). Other research has found evidence to link extremist sympathies to common mental disorders such as depression (Bhui et al. 2019, 6).

Suppose that the process of acquiring an extremist mindset (militant or otherwise) is described as the radicalisation process. Aside from the psychological or other factors that pre-dispose a person to radicalise there is also the question of how the radicalisation process itself works. In truth, there are likely to be many such processes, and multiple different pathways to an extremist mindset if one doesn’t already have it.³⁰ Extremist ideologies reinforce an extremist mindset but one might suppose that such ideologies are only attractive to individuals who have this mindset to begin with. Some accounts of radicalisation see it as something that *happens* to a person, through physical or online contact with extremist ideologues. Other accounts question the assumption that extremism is a ‘communicable disease’ (Wood 2018, 179) to which some people are vulnerable. They see it more as an expression of an individual’s agency, as is suggested by talk of *self*-radicalisation. There are also questions about the role of group dynamics in the radicalisation process, with some influential accounts insisting that extremists who only come into contact with other extremists, and are prevented from interacting with people with different views, are likely to have their extremism reinforced. The resulting ‘crippled epistemology of extremism’ (Hardin 2010) is the result of group dynamics rather than individual choice. It is groups that are ‘the natural habitat of extremism’ (Breton and Dalmazzone 2010, 55).

The jury is still out on whether and how an extremist mindset is acquired. However, regardless of how a person comes to have an extremist mindset, there is the practical question of what, if anything, can be done to counter this mindset. A natural thought is that if having an extremist mindset is partly a matter of how one thinks, then one

way to counter this mindset is to cultivate or inculcate thinking styles that are antithetical to extremist thinking.³¹ Anti-extremist thinking will be realistic rather than utopian or catastrophic. It will respond to conspiracy and apocalyptic thinking with healthy doses of scepticism, humour and irony. As Emcke observes, ‘what is needed is a culture of enlightened doubt and irony – because those genres of thinking are most inimical to the rigorist fanatics and racist dogmatists’ (2019, 111). If extremists or proto-extremists can be trained to ask questions like ‘is that true?’, ‘is there any evidence for that?’, ‘do they know what they are talking about’, and to ask these questions as a matter of course, then it should be possible to counter any extremist tendencies in their thinking. Such questions might also serve as an antidote to the extremist’s pre-occupations with persecution and purity, to the extent that such pre-occupations are baseless.

Just as extremist thinking needs to be countered by antithetical thinking patterns, so the attitudes that underpin extremism need to be countered by antithetical attitudes. Scepticism is again the key. Introducing doubt and self-doubt into the extremist mindset is a way to undermine its excessive certainty and uncompromising attitude. Extremists need, somehow, to be made comfortable with difference, ambiguity and uncertainty. Uncertainty is, in turn, a cure for hate if Emcke is right about hating requiring absolute certainty. Finally, greater open-mindedness, if such a thing can be taught, is the obvious antidote to the extremists’ closed-mindedness and dogmatism. Above all, their othering tendencies need to be countered by helping them to see people who do not share their outlook as human beings who are not to be killed or tortured in the name of some supposed greater good.

Many of these antidotes to an extremist mindset are examples of intellectual or moral virtues. Talk of ‘virtue’ is helpful in this context for reasons that were set out some years ago by Philippa Foot. The Aristotelian virtues, Foot points out, are *corrective*, ‘each one standing at a point at which there is some temptation to be resisted or deficiency of motivation to be made good’ (1979, 8). As noted above, the extremist mindset draws on certain natural human tendencies. This evidence indicates that fanatical thinking patterns are ‘somewhat common’ and that the base rate of such thinking in the population at large ‘does not appear to be low’ (Saucier et al. 2009, 267). If this is right then extremism in one form or another is an example of a temptation to be resisted through the cultivation of corrective anti-extremist virtues. There would be no reason to regard scepticism and irony as *virtues* in this sense if extremism in one form or another were not something by which large numbers of people are tempted. Extremism is the disease for which corrective anti-extremist virtues are the antidote.

How is this antidote to be administered? Can open-mindedness be taught? How can a person who is prone to extremist or fanatical

thinking be made comfortable with ambiguity and uncertainty? At least some anti-extremist virtues are intellectual virtues. In a useful discussion, Baehr argues that ‘fostering growth in intellectual virtues should be a central educational aim’ (2014, 107) and outlines seven practical measures for doing this in an educational setting. Not all anti-extremist virtues are character traits, and a number of them – such as the ability to see out-group members as human beings – might more accurately be characterised as moral rather than intellectual virtues. The extent to which they can be fostered in an educational setting remains an open question. It is an empirical question whether the measures described by Baehr are effective. If they are effective then there is hope for the project of countering extremism by education.

This approach to countering extremism has more going for it than some governmental responses. For example, the U.K. government defines extremism as ‘vocal or active opposition to fundamental British values, including democracy, the rule of law, individual liberty and the rule of law’.³² The inadequacy of this definition is perhaps too obvious to need spelling out. Its unfortunate practical consequences have included the placing of an ‘active duty’ on schools to promote fundamental British values, in the vain hope that the promotion of these values may contribute to countering extremism.³³ Yet if extremism is understood as a mindset, with its distinctive pre-occupations, attitudes, emotions and thinking patterns, then extolling the virtues of the rule of law and democracy is unlikely, on its own, to have much impact, beyond fuelling the sense of resentment felt by marginalised individuals and communities. A more constructive approach is needed, and the discussion above suggests that it might prove fruitful to focus on equipping citizens at an early age with a range of virtues that will reduce their susceptibility to extremism. If an extremist mindset is the problem, then tackling that mindset must be part of the solution.

These recommendations are of particular importance today because of the extent to which recognisably extremist pre-occupations, attitudes and thinking patterns have entered the political mainstream. To take just one example, the supposed victimisation of the United Kingdom by the E.U. has been a key pre-occupation of many English supporters of ‘Brexit’, Britain’s exit from the European Union. What O’Toole describes as ‘a genuine national revolution against a phoney oppressor’ (2019, 164) – the E.U. – is very much in keeping with the extremist cult of victimhood. The fantasy of Brexit as a revolt against oppression both creates and exploits a sense of national self-pity. The uncompromising attitude of the more extreme pro-Brexit faction in British politics is explained by its pre-occupation with victimhood, as well as its hankering after the purest form of Brexit – a so-called ‘clean’ Brexit.

The issue here is not whether Brexit is an extremist policy but whether the arguments in its favour deployed by its most committed proponents are expressive of an extremist mindset. This question must be answered in the negative if all extremism is violent or pro-violence. On the whole, violence is not on the Brexit agenda but it is a mistake to stipulate that extremism must be violent. The mindset of the most hard-line supporters of Brexit is an extremist mindset, not in the sense that it is pro-violence but rather in the sense that its pre-occupations, attitudes and styles of thinking are one that will be familiar to anyone who has made a study of this mindset in other, perhaps more familiar contexts. The resulting polarisation of British politics is, again, something that could have been predicted by anyone with even a passing familiarity with the way that extremists operate. Extremism is a spectrum and it is a serious matter if even mainstream political movements are somewhere on this spectrum.

Notes

- 1 The extremism I am concerned with in this paper is *political* extremism. There are, of course, several other varieties.
- 2 On the relationship between the descriptive and evaluative content of the 'extremist' label see Nozick (1997, 299).
- 3 Nozick notes that 'a simple definition of extremism is not really possible' but that there is 'a cluster of features, some more central than others, that constitutes what might be called an extremist syndrome' (1997, 296). In the same way, there is a cluster of features, some more central than others, that constitute an extremist mindset. In the present discussion I don't try to rank the suggested features of an extremist mindset in order of importance.
- 4 The suitably expansive conception of ideology I have in mind is what Raymond Geuss calls 'ideology in the descriptive sense'. This includes, as well as the beliefs of the members of a group, 'the concepts they use, the attitudes and psychological dispositions they exhibit, their motives, desires, values, predilections, works of art, religious rituals, gestures, etc.' (1981, 5). There are many items on this list that help to constitute a person's mindset.
- 5 Labelling someone as an 'extremist' is rarely understood as a way of complimenting them.
- 6 On ISIS, see McCants (2015), Wood (2015) and Wood (2018). On Breivik, see Seierstad (2016).
- 7 See Saucier et al. (2009), Stankov, Saucier and Knežević (2010) and Medvedovic and Knežević (2018).
- 8 A third extremist preoccupation which, for reasons of space, will not be discussed here, is with a mythic or mythologized past. For an account of this preoccupation in relation to fascism, see the opening chapter of Stanley (2018). See, also, Saucier et al. (2009, 261) and O'Toole (2019, 75–109).
- 9 The purity preoccupation is related to what Jonathan Haidt calls the 'sanctity/ degradation foundation' of conservative morality. If Haidt is right about conservatism's preoccupation with 'stain, pollution and purification' (2012, 171) then one might conclude that conservatism is more likely to be associated with extremism than outlooks that do not have this preoccupation.

- 10 As Nozick notes, a key question is how we distinguish the extremist's non-compromising position from a principled one. As he points out, 'even if one has principles and is convinced that they are right, there can be non-authoritarian ways of maintaining them; one can still be willing to listen to and consider counter-arguments' (1997, 297). Those with an extremist mindset are unwilling to listen or consider counter-arguments. This aspect of the extremist mindset is closely related to its closed-mindedness.
- 11 These vices are discussed in much greater detail in Cassam (2019a), especially Chapters 2 and 5.
- 12 Roberts and Wood (2007, 194–195). See, also, the account of dogmatism in Chapter 5 of Cassam (2019a).
- 13 On hate, see Emcke (2019). On fear, see Appadurai (2006). Self-pity is the focus of O'Toole (2019). Another basic extremist emotion is anger, as described in Mishra (2018).
- 14 In particular, there is what Appadurai calls 'fear of small numbers'. See Appadurai (2006).
- 15 Ben-Ghiat is quoted in a *Washington Post* article as describing a cult of victimization as part of the persona of leaders with authoritarian tendencies. The title of the *Post* article, published on 28 September 2019, says it all; 'Staring down impeachment, Trump sees himself as a victim of historical proportions'. A similar point is made by Jason Stanley in Chapter 6 of his book on fascism (Stanley 2018).
- 16 See Cassam (2019a, 177).
- 17 In their account of what they call 'catastrophizing', Saucier et al. note that 'among militant extremists, there may be an obsession with events perceived as catastrophic and a tendency to portray situations as desperate' (2009, 261).
- 18 There are vivid accounts of ISIS's obsession with the apocalypse in McCants (2015), Wood (2015) and Wood (2018).
- 19 As argued in Cassam (2019b).
- 20 See Saucier et al. (2009).
- 21 On the role of pro-violence in militant extremism see Stankov, Saucier and Knežević (2010).
- 22 As Wittgenstein puts it, 'the *questions* we raise and our *doubts* depend on the fact that some propositions are exempt from doubt, are as it were like hinges on which those turn' (1969, 341). The propositions that extremists regard as exempt from doubt are no such thing.
- 23 See Cassam (2019a) for a defence of this approach.
- 24 Wood, Douglas and Sutton (2012).
- 25 This claim is defended in Chapter 6 of Cassam (2019a).
- 26 This controversial view of the suffragettes is defended in Webb (2014). For a contrary view, see the letter by June Purkis published in *The Guardian* on 6 June 2018.
- 27 Webb (2014).
- 28 It follows from this that it is possible to be a terrorist without being an extremist, just as it is possible to be an extremist without a terrorist. As understood here 'extremism' is a mindset. Terrorism is a tactic. Members of the ANC who planned and carried out bomb attacks that predictably killed civilians were terrorists but it is a further question whether their mindset was extremist.
- 29 While acknowledging that he is not a psychologist Nozick speculates in his brief discussion of extremism that 'there is a determinate extremist personality' (1997, 298).

30 As argued in Cassam (2018).

31 See Saucier et al. (2009) for one version of this approach.

32 See, for example, H.M. Government (2015).

33 H.M. Government (2015).

References

- Appadurai, A. (2006), *Fear of Small Numbers: An Essay on the Geography of Anger* (London and Durham: Duke University Press).
- Baehr, J. (2014), 'Educating for Intellectual Virtues: From Theory to Practice', in B. Kotzee (ed.) *Education and the Growth of Knowledge: Perspectives from Social and Virtue Epistemology* (Chichester: Wiley Blackwell): 106–123.
- Berger, J. M. (2018), *Extremism* (Cambridge, MA: The MIT Press).
- Bhui, K. et al. (2019), 'Extremism and Common Mental Illness: Cross Sectional Community Survey of White British and Pakistani Men and Women Living in England', *The British Journal of Psychiatry*, 15: 1–8.
- Breton, A., and Dalmazzone, S. (2010), 'Information Control, Loss of Autonomy, and the Emergence of Political Extremism', in Breton, Galeotti, Salmon, and Wintrobe (2010): 44–66.
- Breton, A., Galeotti, G., Salmon, P., and Wintrobe, R. (2010), *Political Extremism and Rationality* (Cambridge: Cambridge University Press).
- Brons, L. (2015), 'Othering, an Analysis', *Transcience*, 6: 69–90.
- Button, M. E. (2016), *Political Vices* (Oxford: Oxford University Press).
- Cassam, Q. (2018), 'The Epistemology of Terrorism and Radicalisation', *Royal Institute of Philosophy Supplement*, 84: 187–209.
- Cassam, Q. (2019a), *Vices of the Mind: From the Intellectual to the Political* (Oxford: Oxford University Press).
- Cassam, Q. (2019b). *Conspiracy Theories* (Cambridge: Polity Press).
- Emcke, C. (2019), *Against Hate* (Cambridge: Polity Press).
- Foot, P. (1978), 'Virtues and Vices', in P. Foot (ed.) *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford: Basil Blackwell): 1–18.
- Geuss, R. (1981), *The Idea of a Critical Theory: Habermas and the Frankfurt School* (Cambridge: Cambridge University Press).
- H. M. Government (2015), *Revised Prevent Duty Guidance for England and Wales* (London: The Stationery Office).
- Haidt, J. (2012) *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (London: Penguin Books).
- Hardin, R. (2010), 'The Crippled Epistemology of Extremism', in Breton, Galeotti, Salmon, and Wintrobe (2010): 3–22.
- Kiernan, B. (2008), *The Pol Pot Regime: Race, Power, and Genocide in Cambodia under the Khmer Rouge, 1975–79*, Third Edition (New Haven, CT: Yale University Press).
- McCants, W. (2015), *The ISIS Apocalypse: The History, Strategy, and Domsday Vision of the Islamic State* (New York: St. Martin's Press).
- Mededovic, J., and Knežević, G. (2018), 'Dark and Peculiar: The Key Features of Militant Extremist Thinking Pattern', *Journal of Individual Differences*, 40: 92–103.
- Mishra, P. (2018), *Age of Anger: A History of the Present* (London: Penguin Books).

- Nozick, R. (1997), 'The Characteristic Features of Extremism', in R. Nozick (ed.) *Socratic Puzzles* (Cambridge, MA: Harvard University Press): 296–299.
- O'Toole, F. (2019), *Heroic Failure: Brexit and the Politics of Pain* (London: Head of Zeus Ltd.).
- Ramsey, F. P. (1931), 'Truth and Probability', in R. B. Braithwaite (ed.) *The Foundations of Mathematics and Other Logical Essays* (London: Kegan Paul, Trench, Trubner & Co., Ltd.): 156–198.
- Roberts, R. C., and Wood, W. J. (2007), *Intellectual Virtues: An Essay in Regulative Epistemology* (Oxford: Oxford University Press).
- Saucier et al. (2009), 'Patterns of Thinking in Militant Extremism', *Perspectives on Psychological Science*, 4: 256–271.
- Scruton, R. (2007), *The Palgrave Macmillan Dictionary of Political Thought*, Third edition (Basingstoke: Palgrave Macmillan).
- Seierstad, A. (2016), *One of Us: The Story of a Massacre in Norway – And Its Aftermath* (New York: Farrar, Straus and Giroux).
- Stankov, L., Saucier, G., and Knežević, G. (2010), 'Militant Extremist Mind-Set: Proviolence, Vile World, and Divine Power', *Psychological Assessment*, 22: 70–86.
- Stanley, J. (2018), *How Fascism Works: The Politics of Us and Them* (New York: Random House).
- Webb, S. (2014), *The Suffragette Bombers: Britain's Forgotten Terrorists* (Barnsley: Pen & Sword Books Ltd.).
- Wintrobe, R. (2010), 'Leadership and Passion in Extremist Politics', in Breton, Galeotti, Salmon, and Wintrobe (2010): 23–43.
- Wittgenstein, L. (1969), *On Certainty*, trans. D. Paul and G. E. M. Anscombe (Oxford: Basil Blackwell).
- Wood, G. (2015), 'What ISIS Really Wants', *The Atlantic*, March 2015 Issue.
- Wood, G. (2018), *The Way of Strangers: Encounters with the Islamic State* (London: Penguin Books).
- Wood, M., Douglas, K., and Sutton, R. (2012), 'Dead and Alive: Belief in Contradictory Conspiracy Theories', *Social Psychological and Personality Science*, 3: 767–773.

6b Commentary from Barend de Rooij & Boudewijn de Bruin

Commentary on Quassim Cassam's 'The Vices and Virtues of Extremism'

BAREND DE ROOIJ AND BOUDEWIJN DE BRUIN

The Reign of Terror in France, the Red Army Faction in Germany, Khmer Rouge, Islamic Jihad – examples of extremism abound. But what is extremism, and how does it differ from fundamentalism, radicalism, fanaticism or terrorism?

Perhaps the most straightforward answer is to take the term at face value: extremist views about a topic are views at the very tails of the distribution of possible views about the topic. But reading Cassam's timely and thought-provoking vice epistemological account of extremism shows that such a definition would be far too simple.

According to the view Cassam defends extremism is – in brief – an epistemically vicious pre-occupation with purity and persecution, often accompanied by feelings of hatred. We gladly take the opportunity to raise a few hopefully constructive questions and comments. Our perspective is policy.

From such a perspective, it may be interesting to start with the observation that the widely embraced *Rome Memorandum on Good Practices for Rehabilitation and Reintegration of Violent Extremist Offenders* specifically recommends the inclusion of 'cognitive skills programmes': 'States could consider developing cognitive programs that assist offenders in defining the issues that pushed them towards violent extremist behaviors in the first place and subsequently in formulating objectives and identifying and implementing solutions.'¹

To teach offenders cognitive skills (or epistemic virtue, for that matter), we need to know why they lack them. The literature on extremism uncovers a harrowing array of factors contributing to a person's extremism, including sexual abuse, domestic violence, alcohol and drugs, or just about everything that creates an environment in which little stands in the way of being justified in believing that one is humiliated, ostracised, or degraded. No wonder that in such environments one adopts the simplistic world view of 'us' versus 'them', of good 'friends' versus

evil 'foes', as policy-oriented criminological research suggests. Some evidence suggests that for individuals in such disadvantaged environments us–them thinking may actually be a quite rational generalisation of real lived experiences with Islamophobia and racial discrimination. This response is likely reinforced by the way 'the West' sees them: 'The world has changed tremendously in the last ten years, and that has affected me a lot personally. I grew up believing I was an Amsterdam girl. But after 9/11 I became 'a Muslim'. I remember well receiving the first call after the attacks from a journalist who wanted to know how I, as a Muslim, felt about what had happened. I was being reduced to a single label: I was no longer simply a town councillor, but 'the Muslim' town councillor. That hurt a lot.'²

Us–them thinking is linked, in this policy literature, to a second element: the experience of injustices, perpetrated by the out-group, the 'them': nobility, capitalists, communists, religious groups.

Following this literature, an extremist is a person who (i) experiences injustice, (ii) attributes this to members of an out-group, (iii) considers available solutions to rectify the injustice and (iv) selects an extreme (typically violent) solution.

Is such a view compatible with Cassam's account in which purity and persecution, rather than us–them thinking and injustice are at the forefront? Experiences of persecution may be the starting point of radicalisation, but are they always? Could the account be relaxed somewhat so as to include a concern with injustice? Is a concern with purity perhaps not ultimately a special case of an unjustifiable representation of the social world in terms of 'us' (good, true, unadulterated) versus 'them' (bad, false, contaminated)?

Being pre-occupied with purity and persecution is, for Cassam, not in itself an indication of an extremist mind. As we saw, such pre-occupation has to be accompanied by epistemic vice, which gave us our clear entrance to the policy literature. Still, however, some questions remain.

We wonder, for instance, how easy a task it is to determine whether epistemic vice should count as an indication of extremism. An unwillingness to compromise, for example, may just be very appropriate in the domain of such important causes as racial justice or religious freedom. From a policy perspective, we may actually want to avoid branding someone as epistemically vicious lest we frustrate attempts at resocialisation and rehabilitation.

Consider the various initiatives aimed at establishing restorative justice. Offenders and victims are brought together to facilitate the giving of forgiveness, or to foster mutual understanding and to reduce the force of us–them thinking. Such initiatives were trialled in Spain (ETA), Italy (*Anni di piombo*), and Northern-Ireland (*Troubles*).³ Similarly, work with extremist Salafi (e.g., at the Brixton Mosque in London) suggests that a highly effective method of deradicalisation involves inviting

charismatic and respected Salafi theologians to point out the errors in extremist interpretations of Islamic sources. In such experiments, one determining success factor is the focus on commonality (a pure reading of the early sources of Islam), not difference – not *vice*. One might see this as an application of the principle of charity.

This leads naturally to our final question. Assume a standard economic approach explaining human behaviour in terms of expected utility maximisation. This model is increasingly valuable to explain the adoption of beliefs as well. It's not perfect, but it at least helps us to ask the question of what 'incentives' people have to adopt (in our case) extremist beliefs. Some people join extremist groups for reasons to do with the ideology. But many have only a very superficial grasp of the ideology. They want to escape from home, desire revenge, are in it for adventure, have romantic reasons, feel attracted to the warmth of the group, or want to do penance for their alleged sins. The economic model presents policy makers with the challenge to create an alternative that is more attractive than extremism: a competitor. We take it that Cassam has successfully shown that for someone to appreciate an alternative as a better option, they have to see things right, and this requires epistemic virtue. This is not an easy task at all, but we believe that vice epistemology is developing in a direction that might help policy makers to design the 'cognitive skills programmes' that the *Rome Memorandum* recommends.

Notes

- 1 Good Practice Number 15. See <https://www.thegctf.org/Portals/1/Documents/Framework%20Documents/2016%20and%20before/GCTF-Rome-Memorandum-ENG.pdf>.
- 2 Marjo Buitelaar, 'Discovering a different me': Discursive positioning in life story telling over time, *Women's Studies International Forum*, Vol. 43, March–April 2014, pp. 30–37, at p. 33, <https://doi.org/10.1016/j.wsif.2013.07.017>, quoting Fatima Elatik, a well-known Moroccan–Dutch administrator. One may recall George W. Bush's statement to the effect that 'Every nation in every region now has a decision to make: Either you are with us or you are with the terrorists.' See https://www.washingtonpost.com/wp-srv/nation/specials/attacked/transcripts/bushaddress_092001.html
- 3 See https://ec.europa.eu/home-affairs/sites/default/files/what-we-do/networks/radicalisation_awareness_network/ran-papers/docs/ran_cons_overnv_pap_restor_just_pcve_vot_10022021_en.pdf.

6c Commentary from Marco Meyer

The social epistemic duties of institutions in preventing extremism

MARCO MEYER

Cassam argues that what defines extremists is a particular kind of mindset. This mindset is constituted by two factors: a pre-occupation with the idea of being persecuted or victimised, and a pre-occupation with ‘purity’, be that along ethnic, religious or ideological lines. These factors explain the extremists’ uncompromising attitude. Extremists, Cassam argues, have a sense of certainty in the rectitude of their mindset which is as great as it is unwarranted. It feeds off the epistemic vices of closed-mindedness and dogmatism. The extremist mindset also comes with distinctive emotions and thinking patterns. Extremists characteristically experience hatred, fear and self-pity. The interplay of a concern with persecution as well as with self-pity leads to thinking styles that are both utopian and catastrophic, as well as both apocalyptic and conspiracist.

Cassam’s goal is to describe the mindset approach to extremism and demonstrate its advantages over competing approaches. I will focus on what role epistemic virtues can play in defusing extremism. Cassam offers some important pointers in his article. Virtues are, as he puts it, an antidote to extremism. How can virtues protect against extremism? Cassam appeals to Philippa Foot’s insight that Aristotelian virtues are corrective. They help us to resist temptation.

This view of virtues raises three questions with respect to their protective power against extremism: What are the temptations that extremists give in to? Who is best placed to address these temptations – the extremist-in-making, or people or institutions other than the extremist? And: Which virtues are effective in resisting these temptations?

The temptations of extremism

I take most of my knowledge about extremism from the UK *Prevent strategy* (HM Government 2011). The report is based on academic

studies, intelligence work and consultation with organisations working to prevent extremism.

The report concludes that people at risk of radicalisation are in search for identity, meaning and community (HM Government 2011, 5.22). If these are the temptations that give rise to extremism, we are all subject to them at several points in our lives. The report also finds that extremism is more prevalent among the young and lower socio-economic and income groups (HM Government 2011, 5.26). Yet only a small percentage of people that fit these descriptions develop an extremist mindset. What explains the difference between most of us and people who become extremists?

Cassam makes the case that differences in the mindsets of people at risk of radicalisation have a lot to answer for. The extremist's mindset, he maintains, is morally, politically and epistemically vicious. Vices are traits that their bearers have a certain degree of responsibility for.

The role of institutions

I agree with Cassam that people at risk of becoming radicalised have a responsibility to develop virtues that protect them against extremism. Yet there is at least one other place to look for an explanation for why some people get radicalised and others don't. Based on research in social movement and social network theory, the *Prevent Strategy* argues that radicalisation is a social process that happens in small groups (HM Government 2011, 5.23). That is consistent with Cassam's claim that groups are the natural habitat of extremism. The *Prevent Strategy* reports that extremism is strongly associated with a perception of discrimination and the experience of racial or religious harassment (HM Government 2011, 5.22). It goes without saying that institutions bear at least some responsibility for protecting their subjects, at risk of radicalisation or not, against discrimination and harassment. Yet I want to focus here on whether these institutions have the responsibility to develop epistemic virtues that protect people against radicalisation.

Social epistemic virtues for institutions

Social epistemic virtues relate to the epistemic environment. We are all reliant on the people around us to attain epistemic goods. Epistemic goods include knowledge and understanding. If finding meaning and community are indeed unmet needs in people at risk of extremism, making sense of the social world seems a particularly pertinent epistemic good. Who we interact with when making sense of our experiences matters for the meaning we attach to these experiences. Yet many of the institutions in which people spend much of their time give them little choice about the company they keep, or how their epistemic environment

is structured. The *Prevent Strategy* finds that radicalisation happens in schools, faith institutions, prisons and on social media platforms.

Since these institutions play a large role in structuring the epistemic environment of people subject to them, they have a responsibility to develop social epistemic virtues that prevent extremism. The virtues required are other-regarding – they benefit people subject to the institution, not the institution itself.

To start with, institutions need to understand the structure of the social networks that institutions create for their subjects, in a way that respects privacy and self-determination. There is evidence that social media platforms drive polarisation, for instance about vaccination (Schmidt et al. 2018). Epistemic social networks are easier to study in an online environment than offline. We should not conclude that social media platforms are the only culprits just because we have less research on institutions like schools, prisons and faith organisations.

Institutions should also make deliberate choices about the structures they set up. Students are exposed to the views of all of their classmates on a controversial topic rather than just their group of friends if teachers debate the topic in class. Social media platforms should take diversity of opinion into account when selecting the news items that they display to their users, not just engagement. At a minimum, institutions should not degrade the epistemic environment in a way that supports radicalisation.

In addition to placing the burden of developing epistemic virtues that protect against extremism, we should place at least as much emphasis on holding institutions responsible for developing the social epistemic virtues that can prevent extremism. These virtues are other-regarding, and include virtues connected to the monitoring of the epistemic networks of their subjects, as well as ameliorating the structures that give rise to epistemic networks amenable to radicalisation.

References

- HM Government (2011) *Prevent Strategy*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/97976/prevent-strategy-review.pdf.
- Schmidt AL, Zollo F, Scala A, et al. (2018) Polarization of the vaccination debate on Facebook. *Vaccine* 36(25): 3606–3612. doi: 10.1016/j.vaccine.2018.05.040.

6d Quassim Cassam's Response to Commentaries

REJOINDER TO BAREND DE ROOIJ AND BOUDEWIJN DE BRUIN AND MARCO MEYER

QUASSIM CASSAM

In their sympathetic and thoughtful remarks, de Rooij and de Bruin propose that an extremist is a person who (i) experiences injustice, (ii) attributes this to members of an out-group, (iii) considers available solutions to rectify the injustice and (iv) selects an extreme (typically violent) solution. I want to start by considering (i). What exactly is it for a person to 'experience injustice'?

On a factive reading, it is not possible for a person to *experience* injustice if there is no injustice for them to experience, just as it is not possible for a person to experience a glorious sunset if there is no sunset for them to experience. Extremists typically have grievances to which their extremism is a response, and these grievances may well include the *perception* that they are victims of injustice. However, it is not at all unusual for the perceived injustice to be non-actual. In these cases, it is only in a *non-factive* sense that extremists can be said to 'experience' injustice. The parallel is with the sense in which a person who hallucinates a glorious sunset is 'experiencing' a glorious sunset. The non-factive reading of 'experience' leaves it open that there is no *actual* injustice to which the extremist is responding.

This raises an important question: in deciding whether to characterise a person as an extremist, is it relevant whether their grievances are genuine? Take the case of so-called 'Incel' extremists, that is, involuntarily celibate men who resort to violence in response to what they see as their oppression by women. Since Incels are not actually oppressed by women, one might be more inclined to see them as extremists than genuine victims of persecution who use extreme methods in pursuit of their objectives.

On reflection, however, there is no justification restricting the use of the label 'extremist' to people whose grievances are not genuine. When people with genuine grievances use extreme methods, they are still extremists. On this account of what might be called *methods extremism*,

extremism consists in the use of extreme methods in pursuit of one's political objectives. It is a further question what counts as an 'extreme' method. There are extreme methods that are not violent and using violence in pursuit of one's objectives does not necessarily make one a methods extremist. It all depends on the circumstances and the nature of the violence used.¹

Methods extremism is different from the psychological or 'mindset' extremism which is the focus of my chapter. However, the two are not unconnected. Insofar as having an extremist mindset consists in being pro-violence, psychological extremists are more likely to be methods extremists. A third type of extremism is ideological. To be an extremist in this sense is to have an extremist ideology. What counts as an extremist ideology is too large a question to be tackled here. What is clear, nonetheless, is that psychological, methods and ideological extremism are the three main forms of extremism, and that the philosophy of extremism needs to focus on the nature of extremism in these three senses and the relationship between them.²

In his remarks, Meyer focuses on an institutional response to extremism. He argues that institutions like schools, prisons, and places of worship have a responsibility for helping people to develop epistemic virtues that protect them against radicalisation. The virtues that Meyer has in mind are *social* epistemic virtues, that is, virtues that relate to the epistemic environment. At a minimum, institutions should not degrade the epistemic environment in a way that leads to radicalisation. What this means in practice for schools is that students should be exposed to a variety of different opinions.

It is an empirical question whether exposure to a variety of different views is an antidote to radicalisation. The evidence that bears on this question is far from encouraging. For example, there is some evidence that exposure to opposing views on social media can *increase* political polarisation.³ Nothing that increases polarisation can be an antidote to radicalisation. People are radicalised by *arguments* and by *narratives*.⁴ For example, those responsible for the 7/7 bombings in London in 2005 were radicalised by their acceptance of a narrative about Western atrocities against Muslims.

People who are radicalised by arguments for extremism need to be presented with compelling counterarguments. People who are radicalised by extremist narratives need to be presented with compelling counternarratives. Compelling counternarratives are truthful, have credible sources, and speak to the grievances (real or imagined) by which extremists are motivated. Narratives with these virtues help extremists to make sense of the world and thereby meet what Meyer describes as 'unmet needs' in people at risk of extremism. Counternarratives must aim to help actual or potential extremists to reframe their understanding of current events and challenge their assumptions.

This is not an easy thing to do. For example, the UK government's *Prevent* strategy recognises that people like the 7/7 bombers believe that the 'West is at war with Islam' and is 'deliberately mistreating Muslims around the world' (2011, 47). Faced with this narrative, *Prevent* has nothing better to offer than the assertion that 'far from being at war with Islam', the West is 'making great efforts to address deprivation, human rights issues and governance in Muslim majority countries' (2011, 48). Given the bloody history of Western military interventions in Muslim majority countries, such assertions are worse than useless at countering radicalisation. Of the virtues of effective counternarratives, truthfulness is the most important and also the one that is most obviously missing from current responses to radicalisation.

Notes

- 1 For further discussion of methods extremism, see Cassam (2022).
- 2 See Cassam (2022) for further discussion.
- 3 See Bail et al. (2018).
- 4 See the account of radicalization narratives in Cassam (2022, Chapter 8). On the notion of a narrative, see Fischer (1987).

References

- Bail, C. et al. (2018), 'Exposure to opposing views on social media can increase political polarization', *Proceedings of the National Academy of Sciences* 115 (37): 9216–9221.
- Cassam, Q. (2022), *Extremism: A Philosophical Analysis* (London and New York: Routledge).
- Fischer, W. R. (1987), *Human Communication as Narration: Towards a Philosophy of Reason, Value, and Action* (Columbia: University of South Carolina Press).
- HM Government (2011), *Prevent Strategy* (The Stationery Office).

7 Expectations of Expertise

Boot-Strapping in Social Epistemology¹

Sanford C. Goldberg

1

There is a clear sense in which the existence of expertise makes it easy for nonexperts to acquire justified belief and knowledge in a given domain: a layperson need only accept what a recognized expert tells her (although we might also think that the layperson needs to appreciate that the expert is speaking from expertise). But is there also a way in which the existence of expertise can make it *harder* for nonexperts to acquire justified belief and knowledge in the relevant domain? Consider this possibility: once a domain has experts, the expectations on *all* inquirers who hope to acquire new knowledge in that domain are enhanced. In this chapter, I defend a qualified version of this idea. My argument will focus on the nonexpert who forms *nontestimonial* beliefs in a domain in which there is widely-recognized expertise. My thesis is that in some of these cases, the existence of expertise can make it harder for the nonexpert to acquire justified belief and knowledge in this way.

Some terminology will prove helpful. Consider a subject who forms their beliefs nontestimonially: rather than taking someone's word for it, they form their beliefs on the basis of their own appreciation of the evidence they have acquired at first hand. In such cases, I will call the belief *autonomously-formed*. When an autonomously-formed belief is justified, I will speak of its justification as *autonomous* justification. I do *not* assume that autonomous justification is entirely free of testimony (including expert testimony) altogether. I simply mean to pick out cases in which S forms a belief that p in a way that does not depend for its justification on any testimony that p. Consider the case of Rex, a non-expert, who observes a plant and forms the belief that the particular plant he is observing is a Sand Dune Willow (*Salix Cordata*). This belief is autonomously-formed and (if justified) autonomously-justified, even though (we can imagine) he came to acquire his perceptual competence at discerning instances of *Salix Cordata* by reliance on books on the plants of the Upper Midwest. Even so, in thinking that *this plant* is a Sand Dune Willow, he is relying on his own judgment on the matter at hand: even if that judgment is informed by his past reliance on expert

opinion (from the book), no expert has attested to the proposition he has judged to be true. This is the sense in which his belief is “autonomously formed”.

It is obvious (I hope) that some autonomously-formed beliefs are justified.² Examples abound. Consider the layperson who has a longstanding interest in botany; the parents who, having raised several children, can reliably discern common maladies in their kids; the person whose culinary principles were developed by years of trial-and-error cooking in his own kitchen; the farmer with views about crop yield and pest control developed by decades of her own careful observations; and the amateur brewer who, though lacking any formal education, has developed rules of thumb over the years for making drinkable fermented beverages. We might also include laypersons’ “folk” beliefs in domains such as physics, biology, weather forecasting, and psychology; the veteran poker player whose sense of the goodness of her hand is not based on probabilistic calculations but rather on her developed (albeit inarticulate) feel for the probabilities; amateur gardeners whose rough and ready generalizations about plant care come from long experience; people whose opinions on nutrition are based on careful generalizations regarding the observed effects of their friends’ diets; and those whose hobbies require of them to have a developed sense of judgment in the domain in question. None of these subjects are experts on the topic on which they are forming beliefs; all of them form beliefs in these domains on the basis of the evidence they acquire themselves, and yet even so all of them appear to be in a position in which to form justified beliefs and knowledge on the matter in question. This is so even though in each case there *is* relevant expertise, where any expert opinions on these matters would often (usually?) be both better-informed and more reliable.³

Two important qualifications are called for. First, for the sorts of non-experts just described, I do not assume that *all* of their autonomously-formed beliefs are justified. Surely that isn’t so. (The amateur brewer sometimes acquires unwarranted views as to what makes good beer; the weekend gardener’s beliefs about how to get his plants to flourish aren’t always well-grounded.) Second, I do not assume that *every* sort of nonexpert has some justified autonomously-formed beliefs. Imagine an anti-vaxxer who forms beliefs about the COVID-19 vaccine in an autonomous fashion⁴; his belief will be unjustified. My claim is only that *some* nonexperts are such that *some* of their autonomously-formed beliefs are justified.

How does the development of expertise in a domain bear on the epistemology of (nonexpert) autonomously-formed belief in that domain? Does it have any effect on the conditions for justified belief?⁵

To the best of my knowledge, this is not a topic that has been taken up in the contemporary epistemology literature. I suspect that most epistemologists would think that the answer is obvious. The obvious answer

is that the existence of expertise in a domain is epistemically relevant in a given case of autonomously-formed belief (in that domain) *only to the extent that* the nonexpert is *aware* of the existence of such expertise (and perhaps aware as well of what prevailing expert opinion is). I will call this the “orthodox” view. The orthodox view is supported by the following assumption, which I will designate as the “Doctrine of Ignorance”:

If a non-expert is non-culpably ignorant of the existence (or prevailing opinions) of relevant experts, then that expertise is “blankly external” to the subject’s epistemic perspective.⁶

I suspect that it is *because* most epistemologists will find the Doctrine of Ignorance plausible, that they will regard the orthodox view as obvious.

My case against orthodoxy will be indirect. I will begin (Section 2) by briefly sketching the orthodox approach. Next, I propose an alternative account on which the mere existence of expertise in a domain has a potential (albeit indirect) bearing on the justification of autonomously-formed belief in that domain, *whether or not* the subject herself is aware of the expertise. In Section 3, two types of consideration will be offered on behalf of this account: particular cases and metaepistemological considerations. In Section 4, I present an alternative model, meant to capture the cases presented in Section 3. In Section 5, I consider how the proposed account can handle various other cases. In Section 6, I will suggest two (by my lights, virtuous) implications of the picture on offer; these reside mainly at the intersection of political philosophy and epistemology. Section 7 concludes.

2

According to the orthodox view, the existence of expertise in a domain is epistemically relevant in a given case of nonexpert belief (in that domain). Such a view is motivated by what the epistemological tradition will regard as a perfectly general point about the scope of the materials on which epistemic assessment supervenes. Consider evidentialism, according to which one’s belief is justified just in case it “fits” one’s total evidence. Evidence *not* in one’s possession is simply irrelevant to the question of whether one’s belief fits one’s total evidence – and so is (by evidentialist lights) irrelevant to justification. So insofar as a nonexpert subject is entirely ignorant of the existence of expertise, neither the fact that such expertise exists nor the facts about specific expert opinion are part of her evidence – and so are not relevant to justification.⁷

The temptation to endorse the orthodox view is not limited to evidentialists. This is because the Doctrine of Ignorance itself is perfectly

general: if the nonexpert is ignorant of the existence of expertise in the domain, then that expertise is “blankly external” to the epistemic perspective of the nonexpert – and so would appear to be epistemically irrelevant to her belief. This is so whether one’s preferred epistemology is internalist or externalist, foundationalist or coherentist, reasons-based or reliabilist. One possible complication arises in cases in which one’s ignorance is “culpable” – where, say, one’s ignorance of expertise was itself the result of one’s having exhibited some sort of epistemic vice (and where one was responsible for having developed that vice).⁸ But we can avoid these complications by stipulating that we are only interested here in cases in which one’s ignorance is non-culpable. With this stipulation made, it can seem nothing more than epistemic commonsense to insist that insofar as a nonexpert is ignorant of the existence of expertise in a given domain, her autonomously-formed beliefs in that domain are unaffected by the fact that relevant expertise exists.

3

I want to call this piece of epistemic “common sense” into question. I will do so using two interlocking sorts of considerations: examples and metaepistemological reflections on these examples. The examples are meant to elicit intuitions about particular verdicts; the metaepistemological reflections are meant to reinforce those verdicts.

My first example involves a single subject who is an expert but who does not, in the case at hand, rely on his own expertise in forming a judgment. While such a subject is not ignorant of the existence of expertise, I will argue that the example nevertheless offers important lessons for the epistemology of autonomously-formed belief more generally. Here is the example:

FOOD SCIENTIST

Roger is an expert food scientist for a large food corporation. It happens that Roger also loves to cook at home. As a result of his love of cooking, he has developed a whole set of rules of thumb in the kitchen: what spices work well together, what tastes can be mixed, and so forth. He recognizes that these rules of thumb, though very reliable, are not quite as reliable as his theoretical knowledge as a food scientist. Their virtue, rather, is that they are much easier, less costly, and less time-consuming to apply. One day, when he is at his job at the food corporation, he is asked whether a given combination will yield a result that a majority of consumers would find delicious. He hasn’t done the experiments yet (they would be very time-consuming), but his rules of thumb strongly indicate that the answer is affirmative. On the basis of his knowledge of the reliability

of his rules of thumb, he responds affirmatively. It turns out, however, that he was wrong, and had he done the relevant experiments, he would have known this.

In what follows I argue that when Roger responded affirmatively to the query, his belief was supported by his total evidence, but that even so, it was not justified at that time.

I begin with the claim that, at the time he responds affirmatively, his belief (that the given combination will yield a result that a majority of consumers would find delicious) is supported by his total evidence. He has evidence of a rule-of-thumb variety supporting his belief, and he is aware that the rules themselves are reliably-formed generalizations. In addition, there was nothing Roger knew, or was justified in believing, at the time that would have led him to predict that the mixture would *not* be one that a majority of consumers would find delicious. He could have discovered this, of course; but doing so would have required time-consuming tests which he did not perform at the time.

One might respond by noting that Roger violated a *known professional responsibility to have done the tests*. But while Roger did indeed violate such a responsibility, this does not establish that his belief is unsupported by his total evidence at the time. Let us grant that the following are part of Roger's total evidence at the point of time at which he originally arrives at the affirmative verdict:

F1 There are further tests that can be done to determine whether the proposed combination will yield a result that a majority of consumers would find delicious.

F2 I [= Roger] have a professional duty to do those tests.

F3 I [= Roger] have not done those tests.

The difficulty is that none of these known facts, whether taken separately or in combination, give Roger any reason to doubt the truth of his current belief (based on his rule-of-thumb evidence) that the proposed combination will yield a result that a majority of consumers would find delicious.

The point at issue can be made by construing (Roger's knowledge of) F1–F3 as evidence Roger has of the existence he lacks.⁹ Insofar as Roger is a food scientist, he was aware both that there is further evidence to be had, and that this evidence would be better (more probative) than the evidence he currently has. This is not to say that he knows *what* that evidence will likely support; what he knows, rather, is that *whatever it supports* will enjoy more epistemic support than his current belief enjoys. Now the fact that he has higher-order¹⁰ evidence that there is further (more probative) evidence available – his knowledge of F1 – is part of his current total evidence, and so is already factored into the

assessment of his belief's justification. We can allow that if his evidence of further evidence (= his knowledge of F1) gives him a reason to think that the further (more probative) evidence would tell against his current belief, his current belief is epistemically weakened, and so may be unjustified. But it is part of the story that he has no reason to think that the further (more probative) evidence will tell against his current belief.

Nor should it be thought that the mere knowledge that there is more probative evidence to be had weakens the support provided by one's evidence for one's current belief.¹¹ Here is a parallel case. I look at my watch (of whose general reliability I have some independent evidence) and it reads 2:30, and on this basis, I come to believe that it is 2:30. At the same time, I am aware that there are three reliable clocks within easy walking distance from where I am, and I know that if I check them now, I will get better evidence than I currently have as to the time. (That further evidence would enable me to rule out alternatives that my current evidence does not allow me to rule out: namely, that my watch is presently running fast or slow.) Still, it is simply not true that my knowledge that I have not checked those other clocks weakens the support my belief (that it is 2:30) currently enjoys – and this, despite the fact that I know that if I *did* check those clocks, I would have better evidence than I currently have. So, too, it would seem, for Roger's belief: it remains supported by his rule-of-thumb evidence, despite his knowledge that there is more probative evidence that he could (and should!) have.

At this point the orthodox epistemologist might think to bite the bullet: perhaps Roger's belief is justified after all. Perhaps the proper thing to say is that, while it was professionally irresponsible for Roger to form a judgment without engaging in the test first, even so, the judgment itself was justified at the time.

In response, I want to identify a cost to be paid by any epistemic theory that treats his affirmative (rule-of-thumb-based) judgment as justified. Here, an insistence on the *justified* verdict threatens to disconnect epistemic assessment from the legitimate expectations we have of one another as epistemic subjects. Simply put, company officials expected Roger's judgment to be based on his expertise; it was proper for them to expect this of him; so, the fact that his initial judgment is false, where he would have discovered this falsity for himself if only he had done the tests that were properly expected of him, suggests that his belief does not enjoy the sort of "happy normative standing" that accompanies ascriptions of justification.

Orthodox epistemologists who would insist that Roger's original judgment was justified might think to explain away any impression to the contrary. Perhaps company officials were entitled to expect from Roger more than a (merely) justified judgment; perhaps they were entitled to expect a judgment informed by his own expertise (based on the procedures such expertise calls for). In response, it is uncontroversial that officials were entitled to expect a judgment based on Roger's expertise;

the key question is whether this fact ought to be reflected in our theory of justification.

One way to argue that this fact *should* be so reflected is by appeal to pragmatic encroachment. Such an argument might see FOOD SCIENTIST as yet another example on which stakes drive up the standards of justification. But I think a pragmatic encroachment analysis misdiagnoses what is driving the “not justified” verdict. In particular, I submit that the same verdict holds in *any* case in which Roger is in his role as food scientist – whether the stakes are great (the company is thinking of investing millions into the item) or small (his boss is wondering whether the item is tasty, but not much more hangs on it). In the latter case, an unqualified affirmative judgment based on his rules-of-thumb will still be unjustified; if Roger wanted to enter an opinion in such a context, it ought to be qualified.

But even as I reject pragmatic encroachment, I continue to find it plausible that the theory of justification ought to reflect the legitimate expectations company officials have of Roger. Consider how such officials would respond to the allegation that Roger’s original judgment was justified at the time he responded affirmatively to the query. After scoffing they would offer a dismissive retort:

In that case, we don’t give a hoot about what you call ‘justification’; what we wanted to know was whether Roger’s judgment met with the intellectual standards he was responsible for having lived up to (that’s why we rely on him), and it didn’t.

My proposal is that we should heed their call; we should see in their complaint a brief against orthodoxy and an insight about justification. More specifically, Roger’s judgment is unjustified on the basis of evidence that *he doesn’t have* but which *he was properly expected to have had*. Now I recognize that this metaepistemological consideration in defense of the *not justified* verdict in FOOD SCIENTIST is nowhere near decisive. Still, I hope it can be granted to have *some* force, even if not enough to dislodge the tradition-minded epistemologist’s commitment to orthodoxy.

In FOOD SCIENTIST, the subject himself, Roger, actually had the expertise in question, but owing to the costs (in cognitive effort and time) of forming a belief or judgment based on that expertise, he opted instead to do so on the basis of his hard-won rules-of-thumb. But there are cases in which the subject herself does *not* have the expertise in question, even as she is aware that it exists. Here is such a case:

PARENT’S DIAGNOSIS

Like many parents, Saul has developed a parent’s ability to diagnose common health conditions in his children. So when his daughter Nita comes down with what he takes to be flu-like symptoms, he

comes to believe that she has the flu. While her condition lasts longer than any other flu Saul himself has previously observed, he is heartened by recalling someone having told him of a flu whose symptoms lasted for 6 weeks. (This turns out to be false, but the person seemed authoritative, and Saul had no reason to doubt the testimony at the time.) What is more, Saul has never heard of any other more serious conditions whose symptoms mirror those of the common flu. So he persists in his belief that Nita has the flu. Unfortunately, she has a more serious condition, and had Saul gone to the clinic their pediatrician would have properly diagnosed Nita's condition.

Here it seems that at some point in the course of his daughter's condition Saul's belief to the effect that she has the flu is epistemically deficient. Once again it seems that this epistemic deficiency in his belief is owed (not to the evidence he does have, but rather) to evidence he *doesn't* have. In this case, the evidence in question would be obtained by eliciting the testimony of his family's pediatrician. What I want to say about PARENT'S DIAGNOSIS parallels what I said in FOOD SCIENTIST: the subject's belief is unjustified, despite the fact that his total evidence supports his belief.

I begin with the claim that Saul's belief fits his total evidence. I am stipulating, as part of the story, that none of the testimony he has received to date has ever given him a reason to be suspicious in this case. On the contrary, that background testimony gives him the basis for thinking that his daughter's condition is a particularly long-lasting flu. Second, while Saul is aware that the pediatrician is in a better position to diagnose his daughter's condition than he is, this belief is not sufficient, by itself, to constitute a defeater. The argument for this mirrors the parallel argument in FOOD SCIENTIST. We can reinforce this by showing how the move to regard Saul's awareness of pediatric expertise as a defeater would result in an implausible form of skepticism. For surely Saul need not consult the pediatrician before he counts as knowing that his children have a common cold, or an upset stomach from having eaten too much cake. The point is familiar: the phenomenon of defeat obtains only when one has positive reasons to doubt either the truth of one's belief or the probity of one's basis for that belief and mere awareness of the existence of relevant expertise is not by itself a reason to doubt either of these.

The real issue raised by PARENT'S DIAGNOSIS, then, is whether Saul's belief is unjustified. Here I would reiterate what I argued in the case of FOOD SCIENTIST: an insistence on the *justified* verdict threatens to disconnect epistemic assessment from the legitimate expectations we have of one another as epistemic subjects. Only here the expectations concern the sort of care a parent will provide for his child, under conditions in which access to healthcare is available. There is more to be said in defense of this, but I will postpone further discussion until the next section.

Both FOOD SCIENTIST and PARENT'S DIAGNOSIS involve subjects who are aware of the existence of relevant expertise. But I think that there can be cases in which the subject isn't even aware of relevant expertise, but because of the community in which she resides she should be aware of this. Here is one example:

UPSTATE FARMER

Melissa lived Upstate where she had run her own farm for over 25 years. Recently she moved to the Downstate farming community. On moving Downstate she immediately joins the local farmers' cooperative, which requires all farmers to conform to a strict set of guidelines in their farming practices. The guidelines are given in a 100-page document, and all farmers are required to sign on joining the cooperative. Melissa signed it and read most of it, but she did not read the fine print. (Instead, she assumed – not without evidence – that her fellow farmers would let her know if there were any unusual requirements hiding in the fine print, seeing as how they always talked about such requirements amongst each other.) One day, a fellow farmer asks her what course of treatment on the market was most effective in the fight against the Lesser Cornstalk Borer (a local crop pest). It just so happens that during the last several years she spent on her Upstate farm she had the opportunity to observe the effects of the various courses of treatment on the Lesser Cornstalk Borer. On this basis she had come to the conclusion that course of treatment X is most effective, and so she tells her colleague as much. Unfortunately for Melissa, the fine print of the document she had signed required that all farmers in their cooperative consult with the Downstate Farm Association's advice on courses of treatment for familiar pests (as this advice was based on the advice from the Extension Office of Ag State U, which had conducted extensive trials). Had she consulted with the Extension Office, she would have known that distinct course of treatment Y is the most effective against the Lesser Cornstalk Borer.

I submit that Melissa's belief is epistemically deficient in ways that are reminiscent of the beliefs in FOOD SCIENTIST and PARENT'S DIAGNOSIS: her belief fits her total evidence, but still, owing to evidence she should have had, her belief is not justified. Only here she is not even aware of the expertise in question. If this is so, we have a case in which a belief is unjustified on the basis of evidence the subject didn't have, where the subject *wasn't even aware of the existence of this evidence in the first place*.

Since it should be uncontroversial that Melissa's belief fits her evidence, I will focus on the claim that (despite this) it is not justified. On this matter, several things can be said. First, having joined the Downstate

farmer's collective, Melissa is now properly expected to follow their norms and standards. The fact that she is unaware of this standard does not undermine this expectation. ("Ignorance of the law is no excuse".) To be sure, we might want to say that she has at least a partial excuse for failing to follow their norms and standards: she has what otherwise would have been a justified belief as to the best course of treatment, and given her total evidence she had no reasons for doubting her own belief on this matter. Still, at best these considerations provide her with a partial excuse for believing as she does; they do not provide her belief with a justification. Her fellow farmers in the Downstate Farm Association expect each other to consult with the Downstate Farm Association's expert recommendations; this expectation is legitimate; and yet she failed to do so. One might opt to deny that her failure to do so has any bearing on the justification of her belief, choosing instead to say that (while her belief is justified) she is not to be relied on because of her failing to conform to the local norms. But it seems to me that such an analysis, while possible, leaves epistemic assessment unhappily disconnected from our legitimate expectations of one another as epistemic subjects. It may be that we are forced to accept such an unhappy analysis; but we should do so only if there is no alternative, better account on offer. In the next section, I argue that there is such an alternative.

4

Here is where we stand. There are cases in which nonexpert autonomously-formed belief is based on evidence that would otherwise be sufficient for justification, but where, owing to available expertise which the nonexpert fails to consult, the belief is rendered epistemically deficient. I argued above that the sort of epistemic deficiency we observe in these cases ought to be represented as a *lack of epistemic justification*. If so, evidence one *doesn't* have can defeat the justification of one's beliefs. Still, it is not clear how to model this situation: I have argued that neither the fact that relevant expertise exists nor the subject's awareness of this fact constitutes a defeater. This leaves us with a question: under what conditions (and in virtue of what) is a nonexpert's autonomously-formed belief on a topic on which there is relevant expertise defeated?

I propose to address this matter by appealing to the doctrine of *normative defeat*.¹² Suppose the following conditions hold:

- 1 at time *t* *S* believes that *p*, and *S*'s total evidence is *E*;
- 2 *E* renders *p* propositionally justified;
- 3 at *t* there is additional evidence *E** which *S* does not have, but which she *ought* to have had;
- 4 *p* is *not* propositionally justified on the combination of *E* and *E**.

Taken together, conditions (1)–(4) constitute what I will call *the conditions on normative defeat*. When (1)–(4) are satisfied, E^* contains a *normative defeater* of the propositional justification otherwise enjoyed by S 's belief that p . So understood, normative defeat is the phenomenon whereby evidence one doesn't have defeats the (propositional) justification of one's belief.

The doctrine of normative defeat is premised on being able to make sense of the idea that there is evidence one should have had. For this reason, theories that embrace the phenomenon of normative defeat must confront two fundamental questions. What determines the scope of the evidence one should have had, and what is the source of the "should"? Goldberg (2017, 2018) argued that the "should" has its source in the normative expectations others are entitled to have of one's epistemic condition, whether merely in virtue of one's status as an epistemic subject or else in virtue of the (professional, familial, etc.) role(s) one plays in social practices. The evidence one should have, then, is the evidence one would have if one were to fulfill all of the legitimate normative expectations others have of one's epistemic condition.

The doctrine of normative defeat is controversial. Rather than defending it (for which see Goldberg 2017, 2018), I want to argue that this doctrine will enable us to discern the defeating conditions regarding the justification of an autonomously-formed belief on a topic on which there is expertise. (This result might be regarded as further reason to take this doctrine seriously.) Given an autonomously-formed belief that p in a domain in which there is relevant expertise, this belief's autonomous justification is *normatively defeated* just in case

- i others were entitled to normative expectations of one's epistemic condition, where these expectations are relevant to the belief that p ;
- ii if one had fulfilled all of those expectations one would have had evidence E^* ; and
- iii the combination of E^* and one's current total evidence renders p propositionally unjustified.

If this is correct, it yields a picture on which the existence of relevant expertise has an indirect epistemic significance, in that it potentially exposes autonomously-formed belief to the prospect of normative defeat. We can see how this proposal works by returning to the three examples above.

In FOOD SCIENTIST, Roger is a food scientist employed by a company. On all matters pertaining to his job, he is expected (by his employers and fellow employees) to follow the standards of the food science industry, where relevant. These standards include performing relevant tests. Had he done so, the evidence he would have gotten, in the form of the propositions that accurately capture the results of the test he should have performed,

would have rendered his belief propositionally unjustified. As a result, while it is true that the total evidence Roger currently has is/would be sufficient to justify his belief, that justification is normatively defeated.

In PARENT'S DIAGNOSIS, Saul, *qua* parent, is expected to care for his children and look after their well-being.¹³ These expectations render him responsible for his children's health, and so include expectations to consult with doctors as appropriate. Insofar as this was a case in which a duty of care made it appropriate to have done so, Saul's failure to do so exposed him to the risk of normative defeat. This risk materialized since the evidence he should have had (the testimony of the pediatrician) bears negatively against his belief.

Finally, in UPSTATE FARMER, Melissa was expected to consult with the Downstate Farm Association's policies. Had she done so, she would have learned that the Downstate Farm Association makes recommendations on the treatment of local pests. Had she consulted with these recommendations, she would have learned that the most effective course of treatment in connection with the Lesser Cornstalk Borer is Y. If we add this information to Melissa's total evidence, her belief that the most effective course of treatment in connection with the Lesser Cornstalk Borer is X would no longer be justified. The justification of Melissa's belief is normatively defeated.

In addition to classifying the foregoing cases in a satisfying manner, the proposal handles a variety of other cases well.

In this light, consider a case in which relevant expertise would contradict one's own autonomously-formed belief, yet where intuitively this does *not* bear against the justification of that belief. Here is an example:

CHES

Gideon is an amateur chess player. Despite his amateur status, he has played three or four games a day over the past several years, he has studied various books on chess, and he is currently studying under a chess master. As a result, he is getting very good at chess. His sense for the game has improved dramatically, and his competence at judging for himself the relative goodness of available moves is increasingly reliable. At a certain point in a certain game, he makes a given move, confident in his judgment that there was no better alternative move available to him at the time. However, unbeknownst to him, he had been perfectly set up for a move that would have enabled him to initiate the endgame known as the *Réti manoeuvre*. What is more, this manoeuvre is familiar to Grand Masters; had Gideon consulted with a Grand Master, he would have been told that the move he made was not the best available one.

Intuitively, this is a case in which Gideon's belief (to the effect that he made the best move) might remain justified, despite the fact that it is

inconsistent with expert (= Grand Master) opinion on the matter. In this way the CHESS does not pattern like the other cases we have considered so far: the relevant fact of expertise – the fact that Grand Masters would have recognized the opportunity for the Réti maneuver – does *not* bear on the justificatory status of Gideon's belief.

What is the difference between CHESS, on the one hand, and the cases above, on the other, where relevant expertise *does* defeat the justification of the subject's autonomously-formed belief? The notion of normative defeat characterized above suggests a straightforward answer: CHESS is not a case in which there are others who are entitled to normative expectations of Gideon in connection with his belief. Condition (i) on normative defeat does not hold.

We can reinforce that this is the proper analysis of CHESS with another example. This one involves an amateur car mechanic:

AMATEUR MECHANIC

Samantha is an amateur car mechanic. She loves to diagnose her own car's troubles, and she fixes the smaller problems herself. When problems are minor she is highly reliable in her diagnoses. And she has a good sense of when a problem is not minor; in those cases she takes her car to a professional mechanic. One day, she diagnoses her car with a minor problem, and so forms the corresponding belief. However, if she had consulted with a professional mechanic, she would have learned that the problem, though minor, is not what she thought it was (one of the rare cases in which she was wrong about a minor problem).

Intuitively, given her highly reliable competence at discerning minor car problems, Samantha remains justified in her belief as to the minor difficulty she is having, despite the fact that a professional mechanic would have disabused her of this belief had she consulted with them. In other words, this case patterns with CHESS, and not with FOOD SCIENTIST, PARENT'S DIAGNOSIS, OR UPSTATE FARMER. The best explanation for this, I submit, is that in AMATEUR MECHANIC there is no one who is entitled to expectations of Samantha's epistemic condition regarding her car's problem. The correctness of this diagnosis can be reinforced by considering a variant on this case involving a professional mechanic who was hired to do work on another person's car: if the mechanic were to go with (normally reliable) gut instinct and fail to perform what best practice regards as the proper tests, we would not have the same opinion as to the justifiedness of the belief. (Rather, the case would then pattern as the mechanics' analogue of FOOD SCIENTIST.) This ought to give us some confidence that normative defeat turns on whether there are others who are entitled to normative expectations as to one's epistemic condition.

The picture on offer, then, is this. Relevant expertise in a domain can affect the justificatory status of autonomously-formed belief in that domain. It does so when (i) there are legitimate normative expectations that bear on the subject's epistemic condition in connection with the autonomously-formed belief, (ii) the satisfaction of these expectations would require the belief to be based on expert opinion, and (iii) prevailing expert opinion on the matter clashes with the subject's own autonomously-formed belief. Such a picture embraces the idea that relevant expertise *can* undermine one's justification. But it also recognizes that the mere existence of relevant expertise, by itself, does not do so. This is important for two reasons. First, it enables us to acknowledge that the phenomenon of *robust autonomously-formed justified belief* can persist in a given domain involving expertise, and even when expertise clashes with the autonomously-formed belief itself. (This is illustrated in CHESS and AMATEUR MECHANIC.) Second, it enables us to avoid a common error in domains involving expertise, which is to treat non-expert autonomously-formed belief as somehow epistemically suspect *as soon as expertise develops in the domain in question*. This error, which amounts to an injustice of sorts,¹⁴ is the topic of the next section.

5

So far, I have discussed two types of case: cases in which the subject's failure to get an expert opinion on a matter defeats the justification of her autonomously-formed belief (FOOD SCIENTIST, PARENT'S DIAGNOSIS, and UPSTATE FARMER), and cases in which the subject's autonomously-formed belief remains justified despite the existence of contradicting expert opinion that she did not consult (CHESS and AMATEUR MECHANIC). In both types of cases, I have argued that the proposed account does well. I now want to move on to the third type of case. In it, a subject's autonomously-formed belief is based on good evidence, where expert opinion would only offer *further confirmation* of that belief. Such cases are interesting to me in part because they highlight the possibility of a distinctive sort of injustice – as when such beliefs are regarded by community members as unjustified merely in virtue of the fact that they are not informed by expert opinion.

Let me start with some examples, modeled on cases from the anthropology and philosophy of science literatures. Each involves what we might call “folk traditions” and “folk theory” which persist despite the development of relevant institutional expertise.¹⁵ Here are three vignettes modeled very loosely on actual cases:

CROP VARIETY¹⁶

Zawadi is a family farmer, the fifth generation in her family to farm in the area. She is the beneficiary of the received farming customs

and traditions of the farms in her area. Coming to her as “farming lore,” these practices and procedures are themselves the result of a good number of (informal) experiments by farmers in her area, past and present. (It is not uncommon for individual farms to have up to two dozen fields on their single farm, allowing for a variety of informal experiments; and in addition there is regular interaction with farmers from nearby villages as well, where they exchange ideas about best practice.) Farmers there proudly pass this lore from generation to generation. Zawadi herself reliably follows the lore. When the agricultural experts from the city come to town and observe Zawadi’s practices, however, they are immediately dubious of the reliability of the lore, even as they know of no controlled experiments that cast specific doubts as to her views.

ANIMAL HUSBANDRY

Sonam lives as a subsistence farmer in rural India. As the generations before her had done, so she too follows local farming traditions. These include various animal husbandry practices. Sonam is particularly keen in caring for her several water buffaloes; these she uses for ploughing and pulling other heavy equipment. In the nearby towns, however, there is nothing but scorn for these practices, given that the traditional ways are typically not informed by the results of (institutionalized) scientific animal husbandry.

SHEEPHERDERS¹⁷

Lucas is a shepherd in the English countryside and, like many in the area, comes from a family whose members have done the same for as far back as anyone can remember. His family has several Border Collies who help him in his daily routines, and he inherited a series of practices and protocols from his family regarding the herding of and caring for his flock. Given his renown in his town, he is invited to give a talk at a local University; the audience is polite but skeptical of his rural ways, confident as they are that agricultural science has surpassed local traditions.

Though schematic, these sort of examples illustrate an important point: the development of institutionalized expertise can bring with it a skepticism towards any autonomously-formed beliefs in the domain.

It would be too easy – and it would betoken a facile sort of romanticism – to defend local customs and traditions wherever they are found. Local custom is often the proper target of institutionalized expert criticism; it can involve prejudice and closed-mindedness, and it can reflect rigid local hierarchies that prevent real experimentation and the epistemic goods associated with it. But if romanticism is one pitfall to avoid, so too is a dogmatic form of skepticism. In particular, we should

not disdain (the beliefs that inform) local customs and traditions whenever these operate in domains in which there is a more systematic and institutionalized sort of expertise available.

Happily, the proposal above – to regard (1)–(4) as the conditions on normative defeat – appears to yield the right epistemic verdicts in such cases. On the one hand, given an autonomously-formed belief based on evidence that justifies the belief, the mere existence of relevant institutionalized expertise does not affect that justification. So the fact of relevant institutionalized expertise, by itself, is not a candidate defeater for autonomously-formed belief in that domain. On the other hand, when there is relevant institutionalized expertise whose well-confirmed opinions contradict autonomously-formed belief in that domain, then we have a potential case of normative defeat. Whether this potential is *actualized* – whether the belief in question does suffer from normative defeat – turns on the legitimate normative expectations on the subject’s epistemic condition, and on the content of the relevant expert opinion.

I submit that this is the proper way to assess autonomously-formed beliefs that reflect local tradition and local theory. When these are based on “local expertise” – traditions that gave rise to a systematic body of information and know-how that is warranted on the basis of observation, testing, and well-confirmed empirical theory – we can allow that these autonomously-formed beliefs are *prima facie* justified despite the existence of institutionalized expertise. But even if they are *prima facie* justified, this justification is susceptible to the prospect of normative defeat when the beliefs themselves do not cohere with institutionalized expert opinion. Whether they *are* normatively defeated depends on the prevailing normative expectations others are entitled to have of the believers themselves.

One implication worth highlighting here has to do with the potential for a sort of injustice against those whose autonomously-formed beliefs reflect tradition. Consider cases in which these beliefs are summarily downgraded merely for failing to be based on existing institutionalized expertise. Such a downgrade seems to be both epistemically unwarranted and unfair to those with such beliefs. It is *epistemically unwarranted*, since the fact that the belief was not informed by relevant institutionalized expertise is not, by itself, a reason to doubt that the belief is true. Such a downgrade is *unfair*, since it amounts to a kind of discriminatory attitude toward the tradition in question (and so discriminates without merit against those whose beliefs reflect that tradition). Here I note that this is unfair even if the traditional practices themselves turn out to have been unreliable – and so even if the tradition-bound beliefs were not even *prima facie* justified to begin with.¹⁸

Happily, the picture on offer does not sanction the injustice-constituting epistemic downgrade, since it does not regard the mere existence of relevant institutionalized expertise as a candidate defeater.

It recommends that autonomously-formed belief in domains with institutionalized expertise ought to be assessed on their merits. This includes the evidence on which they were based, together with the prevailing normative expectations to which members in the community are entitled. Expert opinion itself is relevant to the assessment of autonomously-formed belief only if those expectations demand it. This, I submit, is a happy middle ground.

6

In this penultimate section, I want to offer one final big-picture argument for my proposal to regard (1)–(4) as conditions on normative defeat. This argument has to do with a kind of (to my mind, happy) *social-epistemic boot-strapping* that obtains if this proposal is correct. The basic idea can be brought out as follows. According to this proposal, the existence of institutionalized expertise in a domain is relevant to the assessment of autonomously-formed belief in that domain only when others are entitled to expect that beliefs in that domain be informed by this expertise. When others are so entitled, this puts a kind of “pressure” on everyone in the community to become informed (if they are going to have beliefs in that domain at all). Once it becomes (something approximating) common knowledge that we have such expectations, everyone is on notice that autonomously-formed beliefs are acquired at one’s own risk.¹⁹ In this way, these expectations constitute a mechanism by which epistemic communities can boot-strap their way into a more informed public.

I offer the following brief remarks in the development and defense of this picture.

First, there are constraints on when others are entitled to form such normative expectations in the first place. Goldberg (2017, 2018) defends the idea that our entitlement to such expectations is generated by legitimate social practices (perhaps among other things). Participation in a legitimate social practice entitles other participants to expect that one will conform to the norms of the practice, so when these norms require that one satisfy certain epistemic conditions, one is properly expected to do so. Here I submit that institutionalized expertise is constituted, in part, by a set of social practices – practices involving the testing and continued self-correction of methods and procedures, training and certification, the signaling of expertise and public reliance on experts.²⁰

Second, we might offer the following (highly simplistic and schematic!) how-possible story regarding how social-epistemic boot-strapping works. In The Beginning beliefs are formed by individuals using whatever epistemic materials are available. Some individuals are seen to have practical successes in which their beliefs are thought to figure. Local traditions emerge when others copy these individuals and learn from them. The resulting traditions get disseminated more or less widely.

When the tradition's theories are warranted by their local track record, they constitute "local expertise". What I have been calling *institutionalized* expertise arises when matters become institutionalized: modern scientific methods and procedures are employed, ways of certifying their proper usage are implemented, practices emerge in which all of this can be signaled to the greater public, and so forth. When the existence of institutionalized expertise becomes known, a question can be asked of the persisting local expertise: how well does it cohere with institutionalized expertise? Insofar as institutionalized expertise gains adherents within the community, people will begin to normatively expect others in the community to be informed of the existence of such expertise. Once these expectations acquire a sort of social legitimacy (more on which in a moment), people are then entitled to have these normative expectations of one another. And once these expectations become something that is (or approximates) common knowledge, people will then be on notice: one acquires autonomously-formed beliefs in the relevant domain at one's own risk. Presumably many people will opt to go with the institutionalized expertise (to avoid opening themselves up to the prospect of the downgrade associated with normative defeat).

No doubt, the foregoing picture is crude in the extreme. But I think it is useful nevertheless. It highlights a possible mechanism for social-epistemic boot-strapping to take place. What is more, it illuminates at least one decidedly political dimension of the development of institutionalized expertise: when it comes to such expertise, one is entitled to normative expectations of others' epistemic condition only when these expectations have acquired a sort of social legitimacy. I regard this "acquisition of social legitimacy" as an affair that is political through and through. What is at issue is the legitimacy of a certain sort of demand we might make of one another, to the effect that one becomes sensitive to the existence and scope of the relevant expertise. This is the sort of demand that requires authorization if it is to be proper, and the sort of authorization I have in mind is social. I suspect that this sort of authorization can take various forms: perhaps a sufficient majority of citizens have the normative expectation, and this grants implicit democratic authority to the demand; perhaps the state itself provides the authorization, as with significant matters of public health or safety; perhaps authorization comes through explicit deliberation by relevant community bodies, and no doubt there are other ways as well. What is important is that while the case for imposing such expectations on one another is in part epistemic – institutionalized expert opinion is (typically) highly warranted by the total evidence available, and is (typically) more reliable than autonomously-formed belief in that domain – even so, the epistemic part of the story does not exhaust the case that must be made. Simply put, we must bear in mind the need for the political legitimacy of the demands that would be imposed if people were to be entitled to the expectations themselves.

It is this political dimension, I suspect, that is at issue in the sort of outrage that can attend assessments of autonomously-formed belief.

On the one hand, there is the (righteous!) outrage at anti-vaxxers and (most) conspiracy theorists. Their failure is rightly seen as a matter not merely of epistemology but also of good citizenship. I assume that the normative expectations here (to be informed by the best science) are legitimate. For this reason, those who violate these expectations (anti-vaxxers; conspiracy theorists) are regarded as violating a legitimate demand of good citizenship – thereby “free riding” (and so putting undue burdens) on those who vaccinate. And this demand of good citizenship extends to include those benighted few who endorse the conspiracy theory without having had access to the science itself: we might excuse them, but we regard their beliefs as thoroughly unjustified (even if, *per impossible*, they had no access to the science and were informed by what they had every reason to think was good testimony). While everyone is (politically) entitled to their opinion, the demands of good citizenship require more. These sometimes require knowing of the existence of expertise, and basing one’s belief accordingly.

On the other hand, there is the (equally righteous!) outrage that one can feel when one sees a traditional group disparaged merely in virtue of their tradition-bound beliefs and practices. This was seen in CROP VARIETY, ANIMAL HUSBANDRY, and SHEPHERDERS. My reconstruction of this sense of outrage sees it, too, as informed by the demands of good citizenship. Just as these demands bear on us as believers – in our doxastic lives we ought to satisfy the normative expectations others are entitled to have of us – so too this places requirements on us as *assessors* of others’ beliefs – we ought to base our epistemic assessments on the relevant evidence. Insofar as the mere existence of institutional expertise is not itself a reason to question the truth of an autonomously-formed belief, failure to recognize this is not only an epistemic failure but also a violation of good citizenship as well – a way of not properly respecting other traditions and (by extension) of not properly treating those who participate in those traditions.

7

In this chapter, my focus has been the bearing of expertise on autonomously-formed belief. I have formulated and targeted an orthodox view in epistemology, according to which expertise is relevant to a subject’s autonomously-formed belief only to the degree to which she is aware of the expertise. This view, I argued, leads to an epistemology that detaches epistemic assessment from the legitimate expectations we are entitled to have of one another as epistemic subjects. In its place, I have argued that we should see the existence of institutional expertise as highlighting the possibility of normative defeat. And I have offered an account according to which the justification of one’s autonomously-formed

belief is defeated when (i) others were entitled to a normative expectation of one's epistemic condition in connection with the belief, (ii) the satisfaction of this expectation requires one to base one's belief on expert opinion, and (iii) had one done so one's current belief would not be justified. Two additional selling points of this theory are that it opens up the prospect for a kind of social-epistemic boot-strapping, and it highlights the ineliminably political dimension of the phenomenon of expertise.

Notes

- 1 I want to thank Baron Reed for helpful written comments on an earlier version of this chapter. I also want to thank the other members of the Northwestern Epistemology Reading Group (in addition to Reed, this includes Jennifer Lackey, Carry Osborne, Whitney Lilly, Andrés Abugattas, Alex Papulis, Spencer Paulson, Katherine Pogin, Nate Lauffer, Regina Hurley, and John Beverley); the various members of the Facebook page of the Social Epistemology Network, for their feedback to several related queries I posted there; and Heidi Grasswick and Eric J. Olsson, for their thoughtful engagement with this chapter in their written commentaries, included in this volume.
- 2 For the examples to follow, I thank Mark Alfano, Boaz Miller, Julia Staffel, Steven Hales, Kareem Khalifa, Guy Axtel, Adam Green, Alexander Stingl, and many other members of the Social Epistemology Network Facebook page who responded to my query.
- 3 Mark Alfano has suggested to me (private communication) that we might expect the phenomenon of autonomous justification to arise anywhere in which pattern recognition is possible even though causal structure remains opaque.
- 4 It may well be that most anti-vaxxers aren't like this, as they rely on the testimony of alleged (anti-vaxxer) "experts".
- 5 Here I ignore how the development of expertise bears on our understanding of the semantics of our terms (for which see Goldberg 2009).
- 6 This delightful expression "blankly external" is attributed to John McDowell. I borrow it from Van Cleeve (2004) and Littlejohn (2012).
- 7 I my own thinking about how evidence not in one's possession can nevertheless bear on the epistemic standing of one's belief, I have been inspired by the work of my colleague Jennifer Lackey. See especially Lackey (1999, 2005, 2017, 2018).
- 8 There is an extensive literature on "culpable ignorance", which addresses the conditions under which one's ignorance excuses (= when it is non-culpable). See, e.g., H. Smith (1983, 2011), Moody-Adams (1994), Rosen (2002, 2004), and A. Smith (2005, 2007, 2008, 2010, 2015).
- 9 Compare the discussion that follows with the treatment in Ballantyne (2015) regarding one's knowledge of the existence of evidence one doesn't have.
- 10 I note that this use of "higher-order evidence" is not in keeping with others' usage, on which the expression designates evidence that bears on (i) what one's current evidence is, (ii) what one's current evidence supports, or else (iii) one's competence to assess either (i) or (ii).
- 11 Compare Goldberg (2016).
- 12 The term "normative defeat" was introduced to the literature in Lackey (1999). She herself has utilized this notion in various settings as well; see Lackey (2005, 2017, 2018).

- 13 There are two possible parties that are entitled to such expectations. We might say that *Saul's children* are entitled to these expectations; this is so even if they themselves are unaware of this, and so even if they themselves don't form such expectations. Alternatively, we might say that *the state* is entitled to expect this from parents; though here matters are somewhat complicated given that I have formulated the conditions on these entitlements in terms of what other people are entitled to expect, rather than in terms of what an abstraction such as the state is entitled to expect. I will assume that such complications can be worked out, though I won't bother doing so here.
- 14 I am uncertain as to whether this sort of injustice would count as an epistemic injustice in the sense of Fricker (2007). I am inclined to think not. It is still an injustice, however.
- 15 A word about my use of "expertise" here is in order. As I use it, "expertise" designates a systematic body of information and know-how that is warranted on the basis of observation, testing, and well-confirmed empirical theory. Some local traditions and local theories meet this condition; these I dub "tradition-based expertise". (I will contrast these with the sort of expertise that emerges in the practices of modern science, which I dub "institutional expertise".) When local traditions and local theories *fail* to meet the condition on expertise, I will call them "merely local traditions".
- 16 Based loosely on examples from Hansen (2019).
- 17 Based loosely on examples from Collins and Pinch (2014).
- 18 Compare: it is unfair to downgrade the credibility assigned to a woman's testimony merely in virtue of the fact that she is a woman, and this unfairness remains even if it turns out that her testimony was unwarranted – indeed, even if it was as precisely unwarranted as the sexist took it to be.
- 19 The risk, of course, is that of normative defeat.
- 20 There are remaining issues to be addressed, of course, regarding when social practices are legitimate, but those I leave for another occasion.

References

- Ballantyne, N. 2015: "The Significance of Unpossessed Evidence." *The Philosophical Quarterly* 65(260), 315–335.
- Collins, H. and Pinch, T. 2014: *The Golem at Large: What You Should Know about Technology*. (Cambridge: Cambridge University Press).
- Fricker, M. 2007: *Epistemic Injustice*. (Oxford: Oxford University Press).
- Goldberg, S. 2018: *To the Best of Our Knowledge: Social Expectations and Epistemic Normativity*. (Oxford: Oxford University Press).
- Goldberg, S. 2017: "Should Have Known." *Synthese* 194(8), 2863–2894.
- Goldberg, S. 2016: "On the Epistemic Significance of Evidence You Should Have Had." *Episteme* 13(4), 449–470.
- Goldberg, S. 2009: "Experts, Semantic and Epistemic." *Noûs* 43(4), 581–598.
- Hansen, S. 2019: "Farmers' Experiments and Scientific Methodology." *European Journal for Philosophy of Science* 9(32), 1–23.
- Lackey, J. 2018: "Credibility and the Distribution of Epistemic Goods." In K. McCain (ed.) *Believing in Accordance with the Evidence*. (Springer Verlag).
- Lackey, J. 2017: "Norms of Credibility." *American Philosophical Quarterly* 54(4), 323–338.
- Lackey, J. 2005: "Memory as a Generative Epistemic Source." *Philosophy and Phenomenological Research* 70, 636–658.

- Lackey, J. 1999: "Testimonial Knowledge and Transmission." *The Philosophical Quarterly* 49(197), 471–490.
- Littlejohn, C. 2012: *Justification and the Truth Connection* (Cambridge: Cambridge University Press).
- Moody-Adams, M. 1994: "Culture, Responsibility, and Affected Ignorance." *Ethics* 104, 291–309.
- Rosen, G. 2002: "Culpability and Ignorance." *Proceedings of the Aristotelian Society* 103(1), 61–84.
- Rosen, G. 2004: "Skepticism about Moral Responsibility." *Philosophical Perspectives* 18, 295–313.
- Smith, A. 2005: "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, 236–271.
- Smith, A. 2007: "On Being Responsible and Holding Responsible." *The Journal of Ethics* 11(4), 465–484.
- Smith, A. 2008: "Character, Blameworthiness, and Blame: Comments on George Sher's In Praise of Blame." *Philosophical Studies* 137(1), 31–39.
- Smith, A. 2010: "Who Knew? Responsibility without Awareness." *Social Theory and Practice* 36(3) 515–524.
- Smith, A. 2015: "Responsibility as Answerability." *Inquiry* 58(2), 99–126.
- Smith, H. 1983: "Culpable Ignorance." *Philosophical Review* 92(4), 543–571.
- Smith, H. 2011: "Non-Tracing Cases of Culpable Ignorance." *Criminal Law and Philosophy* 5: 115–146.
- Van Cleeve, J. 2004: "Externalism and Disjunctivism." In Schanz, R. (ed.) *The Externalist Challenge*. (Berlin: De Gruyter), 481–495.

7b Commentary from Heidi Grasswick

Goldberg's "Expectations of Expertise: Boot-Strapping in Social Epistemology"

HEIDI GRASSWICK

Goldberg's "Expectations of Expertise" offers a challenging and interesting provocation against what he calls the orthodox view on expertise. The orthodox view holds that in assessing the status of a nonexpert's autonomously-formed belief, the existence of (outside) expertise is relevant "*only to the extent that the nonexpert is aware of the existence of such expertise*" (203). Goldberg believes the attractiveness of this view is grounded in the plausibility of the "Doctrine of Ignorance" according to which, if a nonexpert is "non-culpably ignorant of the existence (or prevailing opinions) of relevant experts, then that expertise is 'blankly external' to the subject's epistemic perspective" (203). In opposition to the orthodox view, Goldberg offers an indirect argument to the effect that in some cases, the existence of expertise that you are unaware of may very well undercut the epistemic state of being justified in your autonomously-formed belief. Goldberg argues this can happen in cases when others have legitimate normative expectations of us to engage with the relevant expertise yet we do not. As a result, if the expert testimony is such that it would have served as a defeater of our belief, our justification for our autonomously-formed belief can be undercut.

At the outset, I am very sympathetic to Goldberg's view that often others do have legitimate normative expectations that they hold us to as epistemic agents and that failures to satisfy these expectations can bear on the status of our epistemic condition. However, I'm not as convinced as Goldberg seems to be that the "justification" of one's beliefs is the best place to pinpoint where or how these expectations exert their pressure on the epistemic lives of nonexperts. This is in part because I take a more capacious view of epistemic justification, according to which the stringency of justification required to believe "responsibly" depends on what we are trying to do with the belief, and much of what we do with our beliefs involves interacting with others. The expectations others have on our states of justification when we do things with our beliefs shift

depending on what is at stake and what exactly we intend to do with our beliefs (and what others expect us to do with our beliefs). But more to the point, Goldberg's focus on potential defeaters of justification that have their source in expertise an agent is unaware of seems to skip over where the real pressure is coming from: legitimate normative expectations that people engage in responsible *inquiry* before they "do" certain things with their beliefs. Furthermore, though Goldberg is surely correct that judgments of the "legitimate normative expectations" ultimately will have to be political (dependent on what the community adopts as their social expectations), this brings up further questions of how we define the legitimate boundaries of those communities that bear on our epistemic responsibilities. There is a great deal of messiness in the interactions between different communities and their normative expectations of each of us, and this will cause problems for Goldberg's view that those legitimate expectations can undercut one's justification when the socially accepted "institutionalized expertise" has not been consulted.

I first want to use the FOOD SCIENTIST case to identify the pressure-point of others' expectations on my belief formation. Here, the scientist is asked for his judgment regarding what combination of ingredients would be (generally found) delicious. Goldberg casts this example in terms of the professional responsibilities the scientist has to his colleagues when he responds to their request for his judgment: as a professional food scientist, they expect that his judgment will be informed by the necessary experiments, not just his home-kitchen practice-based judgments. He lets them down when he fails in this, substituting instead a belief that is as yet only supported by his home culinary experiences. But an important feature of this case is that the scientific evidence on this particular culinary combination does not yet exist! If the colleagues are upset by the scientist's reported judgment, it will be because they expected him to undertake the appropriate scientific work before he testified to them. That is to say, they had expectations that the scientist would have undertaken the appropriate *inquiry* necessary to support claims about the taste results, it is not just a matter of reaching for the appropriate evidence. From their point of view, if he had simply based his judgment on what he'd learned in his own kitchen, he should have either specified that this is all that he is basing his claim on so far, making it more obvious to his colleagues that the company should probably not go forward with investment into this culinary endeavor without further research (this would be a judgment made in his professional capacity), or he should have reserved judgment on the matter until a decision was made to put the time and effort into the experiments. In this case, the food scientist's expertise is not just a matter of knowing what has thus far been determined in his lab, but it also involves knowing how to create the knowledge needed to address a particular question, in a context where the answer requires a fairly high level of justification

(or a particular kind of justification) in order to make decisions about production. Goldberg uses this case as a warm-up, to motivate further points. But it reveals that when we expect someone to either employ their own expertise in a certain context, or consult with a relevant expert as in the later cases presented, we are expecting them to undertake certain forms of inquiry before being willing to state something authoritatively or make decisions and take action on their beliefs. In the later cases presented that involve nonexperts, that inquiry involves investigating whether there is relevant expertise on the matter and if so, consulting it, and likely engaging with it on some level (though what that amounts to might vary in different circumstances).

Goldberg argues that the cost of allowing one's justification to stand in situations where they have not engaged the expertise that might serve as a defeater is too high, in that we would have to let go of the important sense in which we expect others to reach for (presumably) the best knowledge available in being answerable to us. But crucially, those expectations kick in when we are involved in specific practices of interacting with others and when we are depending on them in some way for outcomes that will be based on their beliefs. Goldberg cites "antivaxxers" as a case of those who ignore what he calls institutionalized expertise (in this case, the science of vaccination), and criticizes them for both an epistemic failing and a failing of good citizenry.¹ Because vaccines involve issues of public health, and those who resist a vaccine potentially put others at risk, Goldman is right to note that a failure of good citizenry is involved, and there are epistemic components to this failure. But this case and others also reveal a somewhat sanguine approach to "institutionalized expertise" throughout Goldberg's chapter, which reveals some of the concerns about the boundaries of our communities that I noted earlier.

Goldberg recognizes the dangers of a position that reifies the "institutionalized expertise" of a given society, at the expense of those who are engaged in more traditional or folk practices that include some forms of reliable belief-formation. He wants to avoid associating his position with a certain form of injustice where those engaged in traditional practices are looked down upon or dismissed simply because they are not using the "institutionalized expertise" of society (which he further defines as "expertise that emerges in the practices of modern science" (footnote 16)). Nevertheless, he seems to suggest that if the institutionalized expertise does actually conflict with the traditional practices of knowledge production, it will then serve as a defeater to the justification provided by their traditional practices, presumably because the weight of it being accepted as "institutionalized expertise" serves as evidence for its results that overrides whatever conflicting evidence derived from a traditional or folk practice. This, of course, rules out the possibility that the "institutionalized expertise" may in some cases have ended up missing

something important (or misrepresented something) that a traditional practice is able to tell us about the world.²

Further, appealing to “institutionalized expertise” as part of the community’s expectations on how we justify our beliefs leaves us with an even larger problem to solve: where are the boundaries of the community that is generating these “legitimate” normative expectations? And even more importantly, how does one negotiate one’s way through multiple communities that may have different normative expectations on us? If we need the answers to these questions before understanding when I’m justified or not in maintaining a belief without having consulted with expertise, or even being aware of it, things are difficult indeed.

Of course, having admitted my sympathies with Goldberg’s position that legitimate expectations are placed on us as knowers show that I have not yet solved this deeper problem of multiple communities either! However, by keeping our attention on the epistemic *actions* others are expecting of us – whether that be more inquiry, or achieving a higher standard of evidence when the stakes are high – and noticing that others’ interests in our justifications are legitimate only insofar as we are making decisions, taking actions, and affecting others in the process goes some distance toward maintaining clarity in exactly where and why there is legitimate epistemic pressure on each of us from others.

Notes

- 1 The language of “antivaxxers” can itself be seen as problematically inaccurate in describing many people’s attitudes and approaches to vaccines. “Vaccine hesitant” is a broader term that covers a wider variety of attitudes toward vaccines, and reveals that many who may be labelled as “antivaxxers” differ from public health officials (and vaccine “supporters”) more in their specific goals and concerns than simply their beliefs. See, for example, the work of Maya Goldenberg. Maya J. Goldenberg, *Vaccine Hesitancy: Public Trust, Expertise, and the War on Science, Science, Values, and the Public* (Pittsburgh, PA: University of Pittsburgh Press, 2021).
- 2 See for example Brian Wynne’s well-known example of the many mistakes made by government scientists who attempted to manage the contamination of the soils place restrictions on the Cumbrian sheep farming industry shortly after the Chernobyl disaster – mistakes made in part due their lack of local knowledge about the land, the soils, and the needs of the sheep, and other potential nearby contaminants. Brian Wynne, “Misunderstood Misunderstanding: Social Identities and Public Uptake of Science,” *Public Understanding of Science* 1, no. 3 (July 1992): 281–304, <https://doi.org/10.1088/0963-6625/1/3/004>.

7c Commentary from Erik J. Olsson

In the first part of his complex and thought-provoking contribution to the present volume, Sandy Goldberg argues that there are cases in which a nonexpert's autonomously-formed belief is based on evidence that would otherwise be sufficient for justification, but where this belief is rendered unjustified by (potential) evidence which the nonexpert fails to take into account. An autonomously-formed belief is, roughly, a belief which the person forms on the basis of her own evidence, rather than relying on expert testimony. I will refer to such beliefs simply as "autonomous beliefs".

Goldberg gives various examples in support of his claim. One involves Roger, a food scientist for a large food corporation, who is also an enthusiastic cook. The rules of thumb he has derived from his cooking experience are very reliable, but not as reliable as the scientific method he masters. One day, Roger is asked whether a given combination of ingredients will yield a result that a majority of consumers will find delicious. Rather than carrying out time-consuming experiments, he relies solely on his rules of thumb, in spite of the fact that much depends on the outcome. Roger concludes that the answers are in the affirmative. However, had he carried out the experiments, he would have reached the opposite conclusion, that the mixture is disgusting. In this case, Goldberg argues, Roger is blameworthy and his belief unjustified.

I have little quarrel with the food scientist example, or indeed with the other examples, Goldberg gives in support of his main claim. Rather, I take them to be plausible examples in which autonomous beliefs are unjustified due to evidence not being taken into account. As we shall see, my potential disagreement with Goldberg lies elsewhere. I say "potential" because this comment is as much a critical assessment of Goldberg's account as it is an attempt to grasp its full meaning and consequences.

What is it, then, that in Goldberg's view makes beliefs unjustified in examples of this kind? To explain this, he invokes what he calls the *doctrine of normative defeat*. Suppose, for a starter, that the following conditions hold:

- 1 at time t S believes that p , and S 's total evidence is E ;
- 2 E renders p propositionally justified;

- 3 at t there is additional evidence E^* which S does not have, but which she ought to have had;
- 4 p is not propositionally justified on the combination of E and E^* .

When (1)–(4) are satisfied, E^* is said to be a *normative defeater* of the propositional justification otherwise enjoyed by S 's belief that p .

Goldberg reports that he has argued, in earlier work, that the source of the “ought” here is in the normative expectations others are entitled to have based on a person's participation in various social practices. Thus, the evidence one ought to have had is the evidence one would have had if one were to fulfill all of the legitimate normative expectations. In the food scientist example, others (the CEO, other employees, the customers, and so on) were entitled to have certain normative expectations regarding the methods Roger uses as an employed food scientist. The evidence Roger ought to have had is, then, the evidence he would have had if he were to fulfill all those expectations.

These considerations suggest the following conditions of normative defeat vis-à-vis one's belief that p :

- i others were entitled to normative expectations of one's epistemic condition, where these expectations are relevant to the belief that p ;
- ii if one had fulfilled all of those expectations one would have had evidence E^* ; and
- iii the combination of E^* and one's current total evidence renders p propositionally unjustified.

The key questions, then, when determining whether a given belief is normatively defeated are, first, whether there was any legitimate expectation that one would take further evidence into account and, second, what the result of so doing would have been.

While I very much appreciate the complexity and importance of the problem Goldberg is addressing, this is the point where I start to have some questions regarding his specific solution.

Consider Norman, a member of the Flat Earth League. The Flat Earth League believes that the earth is flat. Moreover, the norms governing its activities prescribe that no member must form a belief about the earth's shape without the prior consultation of the League's earth-shape experts. One day, a trusted friend questions Norman's belief in the flatness of the earth. Norman decides, for the first time, to satisfy himself that the earth is indeed flat. This decision leads him to consider all the traditional evidence to the contrary: that when ships sailing away they disappear, they do so bottom-first; that while at sea elevated areas of land are visible at a greater distance than less elevated areas; that other planets are spherical which would suggest that the earth is spherical, too, and so on. Norman hopes to pinpoint exactly where these arguments go wrong.

However, rather than identifying obvious flaws in the traditional evidence for the spherical form of the earth, Norman, to his astonishment, finds the evidence quite compelling. In fact, so strong is the impact of the evidence that he finds himself unable to resist the conclusion that he has been wrong all along: the earth, he must concede, is spherical after all and not flat. This new belief is an autonomous belief and the evidence upon which it is based would be sufficient for justification in all normal cases. However, since Norman is a member of the Flat Earth League, others (the other members of the League) were entitled to the normative expectation that Norman would consult the League's earth-shape experts before forming his belief about the shape of the earth, which he didn't. If Norman had fulfilled this expectation, the experts would have told him that the earth is flat. Their testimony in combination with Norman's current total evidence renders his belief that the earth is spherical normatively defeated, on Goldberg's account, and therefore unjustified. Yet, we may assume that Norman has exactly the same evidence that we have for this belief, the only difference between him and us being that he is, in addition, a member of the Flat Earth League, a fact which itself, however, has no bearing on the shape of the earth. Had he not been a member of the League, he too would have been justified in his belief that the earth is a sphere. This, it seems to me, is the wrong result in this case. Surely, Norman is justified in believing that the earth is a sphere in spite of the fact that he is a League member.

Goldberg is of course aware that his theory faces problems of this kind. In the penultimate section of his contribution, he asserts that there are "constraints on when others are entitled to form such normative expectations [regarding the consultation of expertise] in the first place". One proposed constraint is spelled out as follows (my emphasis):

[W]hen it comes to the sort of expectations I am discussing above, the entitlement is generated by *legitimate* social practices. Participation in a *legitimate* social practice entitles other participants to expect that one will conform to the norms of the practice, so when these norms require that one satisfy certain epistemic conditions, one is properly expected to do so. Here I submit that institutionalized expertise is constituted, in part, by a set of social practices – practices involving the testing and continued self-correction of methods and procedures, of training and certification, of the signaling of expertise, and of public reliance on experts.

Hence, in addition to there being legitimate normative expectations on the part of other people, the underlying social practice in which they participate must itself be legitimate for a belief to run the risk of being normatively defeated. The legitimate expectations that may normatively defeat an autonomous belief do not come from any social practices in

which a person participates, only from those practices that are themselves legitimate. The examples Goldberg gives of normative defeat, including the food scientist, are generally of the kind of which it is reasonable to assume that the social practice in question is legitimate.

As for the Flat Earth League, its members have a legitimate normative expectation that Norman complies with the norms of the League of which he is a member. However, from where we stand, the League's social practice fails to be itself legitimate. Therefore, the League members' expectations lack normative force. Specifically, they lack the normative force needed to render Norman's belief unjustified.

Hence, condition (i) in Goldberg's account of normative defeat should be supplemented, or clarified, as follows:

(i') others *participating in a legitimate social practice* were entitled to normative expectations of one's epistemic condition, where these expectations are relevant to the belief that p.

In my view, the account characterized by conditions (i'), (ii), and (iii) stands a good chance of providing a correct account of normative defeat, at least in outline. The account essentially states that one shouldn't rely solely on one's own evidence in cases in which others are entitled to expect that one would consult legitimate evidence. As Goldberg notices, it has the desirable consequence that beliefs deriving from local customs and traditions are not normatively defeated solely in virtue of the fact that there exists superior evidence bearing on the matter deriving from a legitimate social practice. They are not defeated if taking this evidence into account would not make the beliefs in question unjustified (but, say, only further supported).

In fact, even antivaxxers, who Goldberg rarely misses an opportunity to criticize, may very well think that Goldberg has delivered a correct account of normative defeat. It is only that they have very different views about what specific social practices are legitimate. This is so even if we add a general account of what practices or bodies of expertise are legitimate along the lines of what Goldberg writes in a footnote: "I use the general term 'expertise' to designate a systematic body of information and know-how that is warranted on the basis of observation, testing, and well-confirmed empirical theory". Antivaxxers would presumably have different views on what is "warranted" and "well-confirmed".

Generally, even with Goldberg's account of normative defeat in place, there is plenty of room for disagreement about whether or not a social practice or alleged expert is in fact legitimate in a concrete case. We who believe in science, including virology and related disciplines as currently pursued, will be inclined to think that the expertise in this area is legitimate and that the views of anti-vaxxers, who should have consulted the experts but have failed to do so, have thereby been firmly defeated. A (moderate) antivaxxer might respond that, while much science

is reliable, the particular disciplines in question have, regrettably, been compromised by the corporate influence of Big Pharma, the effect being that the putative experts are not *bona fide* exemplars of their kind. Goldberg's account of normative defeat, as I have described it, is not very explicit on which concrete social practices are in fact legitimate, and I think this is as it should be. For it is not part of the very concept of normative defeat, relative to a social practice and a standard of the legitimacy of such practices, that some social practices are legitimate and others not.

This leads me to my final point. Goldberg thinks that his account, if correct, underpins a kind of "social-epistemic bootstrapping", which he thinks is "happy". He explains:

According to this proposal, the existence of institutionalized expertise in a domain is relevant to the assessment of autonomously-formed belief in that domain only when others are entitled to expect that beliefs in that domain be informed by this expertise. When others are so entitled, this puts a kind of "pressure" on everyone in the community to become informed (if they are going to have beliefs in that domain at all). Once it becomes (something approximating) common knowledge that we have such expectations, everyone is on notice that autonomously-formed beliefs are acquired at one's own risk. In this way these expectations constitute a mechanism by which epistemic communities can bootstrap their way into a more informed public.

Yet, the desirability of this sort of bootstrapping depends of course on the *de facto* legitimacy of the institutionalized expertise in question. If the experts appealed to are the real thing, we should surely welcome bootstrapping, which will then indeed result in a more informed public. But then we also have the unfortunate cases in which this is not so – and we know that scientists are no exception to the rule that human beings are susceptible to various psychological biases and sociological forces, whose potentially truth-derailing effects can lead even our best experts astray at any given stage of inquiry, not to mention their less talented or less honest colleagues (although we hope that science will become less dependent on the vagaries of the human mind as it progresses). Bootstrapping in such cases has the unfortunate effect of cementing the errors introduced by defective experts by making it socially more difficult to dissent on the basis of one's own autonomously-formed beliefs. Even so, since most scientific practices are surely reliable, bootstrapping is *mostly* a good thing. My point is that it is not *always* a good thing, and – I would like to add – that we somehow need to safeguard against the cases in which it is not.

7d Sanford C. Goldberg's Response to Commentaries

Responses to Grasswick and Olsson

I want to thank Heidi Grasswick and Erik J. Olsson for their thoughtful engagement with my chapter. In this brief reply, I will respond to each of them in turn.

In her contribution, Grasswick identifies three related concerns with my proposal. First, while she agrees with my big-picture claim asserting the epistemic significance of normative social expectations, she argues that we should not construe such expectations as affecting *doxastic justification*. Second, while I appeal to normative defeat as the mechanism by which such expectations acquire epistemic significance, she thinks we would do better to appeal to the practical stakes that are in play (and in what we *do* with our beliefs). And third, while I characterize the content of these normative social expectations as concerning the subject's epistemic state, she thinks that we do better to treat them as expectations that the subject engages in certain forms of inquiry.

I feel the force of each of Grasswick's concerns. I acknowledge as well that I am far from confident in the soundness of my proposal for modeling the epistemic significance of normative social expectations. Still, I think some things can be said in defense of that proposal, and I will use Grasswick's concerns as an occasion to say a little more.

I begin with my reason for locating the epistemic significance of normative social expectations in the theory of doxastic justification. It should be uncontroversial that doxastic justification is the status a belief has when it was formed and sustained in a way that passes some threshold of goodness with respect to the twin aims of acquiring truth and avoiding falsehood. But when it comes to articulating that threshold, I worry that no particular threshold can be induced from objective criteria such as reliability or accuracy. For this reason, I think that we do better to treat the reliability threshold itself as set by what others are properly entitled to expect of one, *qua* epistemic subject. I tried to articulate and defend this view at length in my (2018). If I am right in my meta-epistemological account of what determines justification, then the subject matter of this chapter – how the existence of expertise in a domain

might affect what is properly expected of nonexperts in that domain¹ – is equivalent to exploring how the existence of expertise in a domain affects *the justification* of belief in that domain.

This meta-epistemological consideration also motivates my responses to Grasswick's second and third concerns. My approach to normative social expectations is part and parcel of a view on which epistemic justification reflects the range of expectations we are entitled to have of one another *qua* epistemic subjects. All parties should agree that these expectations can be affected by the social practices in which a subject participates. By contrast, the bearing of practical stakes on epistemic assessment is more controversial.² I see it as an advantage of my proposal that it doesn't have to take a controversial view on that matter. (Of course, this only means that I prefer the controversial position that treats social expectations themselves as relevant to epistemic justification.) Finally, while I myself am very sympathetic to the idea that normative social expectations include expectations that subjects perform certain inquiries (for which see Goldberg 2017, 2020, forthcoming a, and forthcoming b), I also think that these expectations bear on the subject's current doxastic state as well, and it is the doctrine of normative defeat that connects the expectations of inquiry with those that bear on the epistemic assessment of belief.

In his contribution, Erik J. Olsson rightly notes that in at least two related respects things can be less rosy than I have depicted them. First, while I tend to focus on experts who are epistemically virtuous, there can be epistemic communities that defer to "experts" who are not. Second, while I tend to focus on the benefits of the sort of social-epistemic "bootstrapping" that arises in connection with the expectations associated with expertise, there are cases in which, owing to deference to "experts" who are misguided or simply wrong, the bootstrapping phenomenon can actually make things worse, epistemically speaking; they can "cement ... the errors introduced by defective experts by making it socially more difficult to dissent on the basis of one's own autonomously-formed beliefs".

I think Olsson is correct in both of these allegations, though I would hope to be able to acknowledge these points in a way that is broadly consistent with my proposal. (Since Olsson may well agree that this is possible, this may not be any criticism of his commentary.) Although I did not highlight it in my contribution here, in my (2018) I tried to argue that there are *epistemological* constraints on legitimate normative expectations regarding another's epistemic condition. These constraints reflect the two core aims which give the standards of epistemic assessment their point: the acquisition of truth and the avoidance of error. The degree to which conforming to a set of candidate standards enables a subject to do well in connection with these twin aims is a fully objective matter. Normative social expectations do not affect this. Their role is

rather to articulate (i) what evidence ought to be taken into account in epistemic assessment, and (ii) where the relevant threshold (e.g. of reliability) ought to be drawn. Since baselines for both (i) and (ii) are set by (what we are entitled to expect of one another merely in virtue of) our status as epistemic subjects, any normative social expectation that violates either baseline is *ipso facto* illegitimate. Consider in this light epistemic communities in which there is an expectation to defer to “defective experts”. Given the epistemic constraint on legitimate normative expectations, these communities’ expectations may well run afoul of those constraints, and so may be illegitimate. (See Goldberg 2018 for details.) This also suggests how I would approach Olsson’s worries about the bad epistemic effects of “bootstrapping” in cases in which “experts” go astray. I would hope that the epistemic constraints on legitimate social expectations would provide a principled way to rule out many (if not most or all) of the cases of “bad” bootstrapping, as having no bearing on assessments of doxastic justification. However, this is not something I addressed either in the present chapter or in my (2018), so it remains to be seen whether this hope can be realized. And if it cannot, then Olsson has identified a potentially unhappy implication that will have to be embraced.³ Whether this is so will have to await future work.

Notes

- 1 Here I should correct the record on one point. My claim was *not* that the mere existence of expertise whose content conflicts with “traditional” belief defeats the latter’s justification. There is the further requirement that others be entitled to expect people to conform to expert opinion. Part of the burden of the chapter was to describe cases in which others would be so entitled, as well as cases in which they would not.
- 2 For a defense of the view that stakes affect the justification of *action*, not belief, see Reed (2010, 2012, 2013) and Goldberg (forthcoming a).
- 3 It is worth noting, though, that a claim in this vicinity – that normative social expectations can give rise to unpossessed evidence defeating one’s knowledge, even when the unpossessed evidence is misleading – is an implication that many appear already to embrace. Consider the “newspaper case” raised by Harman (1973: 143–144) or the “unopened letter” case in Pollock (1986: 192).

References

- Goldberg, S. Forthcoming a: “On the Epistemic Significance of Practical Reasons to Inquire.” *Synthese*.
- Goldberg, S. Forthcoming b: “Epistemic Autonomy and the Right to Be Confident.” In J. Matheson and K. Loughheed, eds., *Epistemic Autonomy* (New York: Routledge).
- Goldberg, S. 2020: “Norms of Inquiry in the Theory of Justified Belief.” In N. Ashton, M. Kusch, R. McKenna, and K. Sodoma, eds., *Social Epistemology and Relativism* (New York: Routledge).

- Goldberg, S. 2018: *To the Best of Our Knowledge: Social Expectations and Epistemic Normativity* (Oxford: Oxford University Press).
- Goldberg, S. 2017: "Should Have Known." *Synthese* 194(8): 2863–2894.
- Harman, G. 1973: *Thought* (Princeton: Princeton University Press).
- Pollock, J. 1986: *Contemporary Theories of Knowledge* (Maryland: Rowman and Littlefield).
- Reed, B. 2013: "Practical Matters Do Not Affect Whether You Know." In M. Steup and J. Turri, eds., *Contemporary Debates in Epistemology*, 2nd ed. (London: Wiley Blackwell), pp. 95–106.
- Reed, B. 2012: "Resisting Encroachment." *Philosophy and Phenomenological Research* 85: 465–472.
- Reed, B. 2010: "A Defense of Stable Invariantism." *Noûs* 44: 224–244.

T&F Proofs – Not for Distribution

8 Fake News, Conspiracy Theorizing, and Intellectual Vice

Marco Meyer and Mark Alfano

1 Introduction

Fake news and conspiracy theories spreading over the internet are a major challenge to public debate discourse and even democratic deliberation.¹ How can we address this challenge? Systemic changes are presumably needed that will require regulation, legislation, and industry intervention, but individuals also play a role and may wish to respond. In this chapter, we focus on the dispositions of individuals, as there is some evidence that there are individual differences in the propensity to endorse and spread fake news and conspiracy theories (Guess et al. 2019; Lazer et al. 2018). Our focus is on intellectual virtues and vices. Intellectual virtues are character traits that support their bearers in gaining and spreading knowledge and understanding (Roberts and Wood 2007). Intellectual vices are deficits in intellectual virtue, undermining the ability to gain or transmit knowledge and understanding (Cassam 2018).

We present findings from two survey experiments measuring intellectual virtue and vice, fake news endorsement, and conspiracist beliefs. The first study was exploratory; the second study was confirmatory and pre-registered.² Across two studies, we collected data on the intellectual virtues of nearly 2000 people from the United States, eliciting their intellectual virtues using a validated survey instrument. Analysis shows that intellectually vicious people are more likely to endorse conspiracy theories. This finding supports claims by epistemologists that conspiracy theorists suffer from intellectual vice (Cassam 2016, 2018, 2019).³ Yet the current experiment looks beyond conspiracy theories by showing that endorsement of fake news is also associated with epistemic vice.

Our experiments also make a contribution to vice epistemology. Vice epistemology is the branch of epistemology that concentrates on the nature, identity, and epistemological significance of intellectual vices (Cassam 2016). Quassim Cassam has suggested that some intellectual virtues and vices may be “stealthy”. A trait is stealthy if possessing the trait stands in the way of knowing that you have the trait (Cassam 2015). The

current experiments use a self-assessment approach to measure intellectual virtue and vice, including dispositions like intellectual humility and arrogance that might intuitively seem to be stealthy (Alfano et al. 2017). That this measure is associated with questionable beliefs suggests that intellectual vices are not fully stealthy. People seem to have some knowledge about their intellectual character traits, even when those traits get in the way of other knowledge.

We study epistemic virtue and vice in an online environment in part out of convenience but also because the internet is one of the main breeding grounds of conspiracy theories and fake news. Recent work has found intense conspiracy theorizing on various online platforms, including Reddit (Klein et al. 2018, 2019) and YouTube (Alfano et al. 2018, 2020). Participants for the study were recruited via Amazon Mechanical Turk, an online crowdsourcing platform. This is an appropriate setting to study conspiracy theories and fake news spread over the internet, where the volume, velocity, veracity, and variety of information sources are unique compared to older, more traditional media infrastructures (Alfano and Klein 2019). Our experiments suggest that epistemic virtue appears to influence whether people place trust intelligently or wisely online.

Section 1 introduces the first experiment, describes the sample in terms of its intellectual virtues and their propensity to endorse conspiracy theories and fake news, and shows regression results of the first experiment. Section 2 introduces the second experiment and shows regression results for it. In the final section, we discuss the implications of this research for virtue and vice epistemology. We focus on whether intellectual vices are stealthy and on implications for trust on the internet. We conclude with reflections on directions for future research.

2 Study 1

2.1 *Materials and Methods*

Participants were recruited using Amazon Mechanical Turk. The eligibility criteria were living in the United States and being 18 or older. Respondents were paid \$2 for participation. A total of 1,357 people participated, of which 975 passed an attention check. Participants who failed the attention check were excluded from the analysis. The participants were on average 40 years old ($SD=13$). 52% of the sample was female. 79% of the sample was White/Caucasian, 9% was African American/Black, 5% was Asian or Pacific Islander, and 5% was Hispanic; the remaining 2% were other or did not disclose ethnicity. 53% had obtained a bachelor's or a higher degree. Mean household income was 57,000 USD per year. Table 8.8 in the appendix contains full descriptive statistics.

Each participant answered demographic questions about age, gender, race, education, household income, religion, political affiliation, and news consumption. We measure religiosity by asking respondents how important religion is to them, on a 5-point scale from “not at all important” to “extremely important”. 42% reported to find religion not at all important. The remainder of the sample is roughly evenly split between according religion slight, moderate, strong, and extreme importance.

We measure political affiliation by asking participants whether they “consider themselves a Republican, a Democrat, an Independent, or what?” Responses are “Strongly Democratic”, “Weakly Democratic”, “Independent (Lean toward Democratic party)”, “Independent”, “Independent (Lean toward Republican party)”, “Republican (Weakly Republican)”, “Republican (Strongly Republican)”. 42% consider themselves Democrats, 22% Republicans, and 36% Independent.

We measure news consumption by asking “How often do you get news from the following sources” (5-point scale: “never”, “rarely”, “sometimes”, “often”, “very often”) for the following news sources: printed newspapers, social networks, TV and Radio, Online Newspapers, and News Aggregators. All five news sources are widely used by respondents, with an average between 3.1 (online newspapers) and 3.4 (TV and Radio).

In the following, we discuss the instruments used to measure intellectual virtue, credence placed in conspiracy theories, and fake news endorsement.

2.1.1 *Measuring Intellectual Virtue*

We measured intellectual virtues using a validated survey instrument (Alfano et al. 2017). The scale provides a measure of intellectual humility. Intellectual humility is only one among many intellectual virtues. However, Alfano et al. have worked with an extensive definition of humility. Using 23 items, the scale measures four related virtues: open-mindedness, intellectual modesty, engagement, and corrigibility.

The constructs are defined in Table 8.1. While the four dimensions do not provide a comprehensive measure of intellectual virtue, the measure is broad enough for the purposes of this experiment. Responses were scaled on a 5-point agree-disagree scale.⁴

We calculated scores for individual virtues by taking the average of the relevant items, and transforming the scale to 0–100, for example, 50 corresponds to the scale point “neutral”, and 100 corresponds to “strongly agree”. Figure 8.1 shows the results. On average, respondents score highest on open-mindedness (average 80), followed by engagement (69) and corrigibility (66). Interestingly, people self-report a comparatively low level of modesty, with the median respondent scoring 53, just above the neutral point.

Table 8.1 Overview of Alfano et al.'s intellectual humility scale. Contrary vices are listed in parentheses after the virtues they oppose. Items marked with (R) are reverse-scored.

Intellectual Virtue	Definition	Example Item
Open-mindedness (Intellectual Arrogance)	Acknowledgment of the limitations of one's knowledge, especially relative to others, and a desire to gain knowledge irrespective of status.	I don't take people seriously if they're very different from me. (R)
Intellectual Modesty (Intellectual Vanity)	Low concern for how one's intellect is perceived, and for one's intellectual reputation.	I like to be the smartest person in the room. (R)
Engagement (Boredom)	Motivation to investigate things one doesn't understand, particularly in response to encountering ideas different from one's own.	I enjoy reading about the ideas of different cultures
Corrigibility (Intellectual Fragility)	Resilience in emotional response when confronted with challenges to one's knowledge or intellectual abilities.	I appreciate being corrected when I make a mistake.

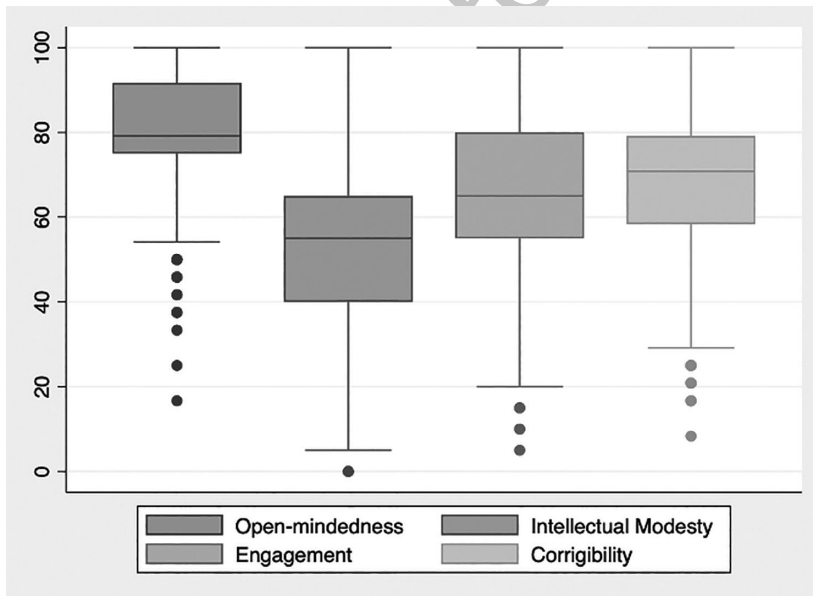


Figure 8.1 Summary of intellectual virtues.

For the initial analysis of the data, we constructed a summary measure of intellectual virtue rather than analysing each virtue separately. We obtained the measure by taking the average scores across all items.

2.1.2 *Measuring Conspiracist Thinking*

Conspiracy theories are explanations for (purported) phenomena that invoke a conspiracy. Sometimes they also include an epistemic component that hypothesizes that the conspirators systematically distort evidence about their activities and even existence. Some conspiracy theories are true (Dentith 2016; Harris 2018). We are interested in conspiracy theories that – for all we know – are false and not inferences to the best explanation (Harman 1965). Some of these conspiracy theories may still turn out to be true, but we maintain, in line with Cassam, that believing such conspiracy theories is a defeasible sign of intellectual vice (Cassam 2016).

To elicit the propensity to endorse conspiracy theories, we used an established measure from political science (Oliver and Wood 2014). Participants were presented with five conspiracy theories in random order and asked whether the statements presented were true or false, on a 5-point scale (“definitely false”, “probably false”, “do not know”, “probably true”, “definitely true”). Table 8.2 presents the items and proportion of respondents endorsing each of the statements. We took respondents to endorse a statement if they replied “true” or “definitely true” to construct the table. More than one-third of respondents endorsed at least one conspiracy theory. For the purpose of further analysis, we take the average of the five items of the score on the 5-point scale described above.

Table 8.2 Measure of conspiracist thinking

<i>Conspiracy</i>	<i>Endorsement (%)</i>
The U.S. invasion of Iraq was not part of a campaign to fight terrorism but was driven by Jews in the United States and Israel.	11
Certain U.S. government officials planned the attacks of September 11, 2001, because they wanted the United States to go to war in the Middle East.	18
President Barack Obama was not really born in the United States and does not have an authentic Hawaiian birth certificate.	13
The financial crisis of 2008/2009 was secretly orchestrated by a small group of Wall Street bankers to extend the power of the Federal Reserve and further their control of the world’s economy.	19
Billionaire George Soros is behind a hidden plot to destabilize the American government, take control of the media, and put the world under his control.	16

2.1.3 Measuring Fake News Endorsement

Fake news refers to content that presents (typically) false or misleading claims as news (Gelfert 2018; Lazer et al. 2018). Fake news can be spread for many different reasons, from pure trolling to driving advertisement revenue to corporate or state-led disinformation campaigns.

We developed a new instrument to elicit the propensity of respondents to endorse fake news. Each respondent was presented with ten screenshots of articles from news and fake news websites in random order. Participants were asked whether the article displayed was credible, on a 5-point scale (“strongly disagree”, “disagree”, “neutral”, “agree”, “strongly agree”).

Figure 8.2 shows the proportion of fake news items people find credible. To construct the graph, we took respondents to find an article credible if they replied “agree” or “strongly agree” to the question of whether the article was credible. Almost four of five respondents found at least one of the fake news items presented to them credible. On average, respondents found 1.5 out of 5 fake news articles credible. For the purposes of further analysis, we work with the mean across all five fake news items of the 5-point scale described above.

It is noteworthy that endorsement of conspiracy theories and endorsement of fake news are correlated, with a coefficient of 0.6. This result suggests that there might be an underlying factor explaining both types of pernicious beliefs.

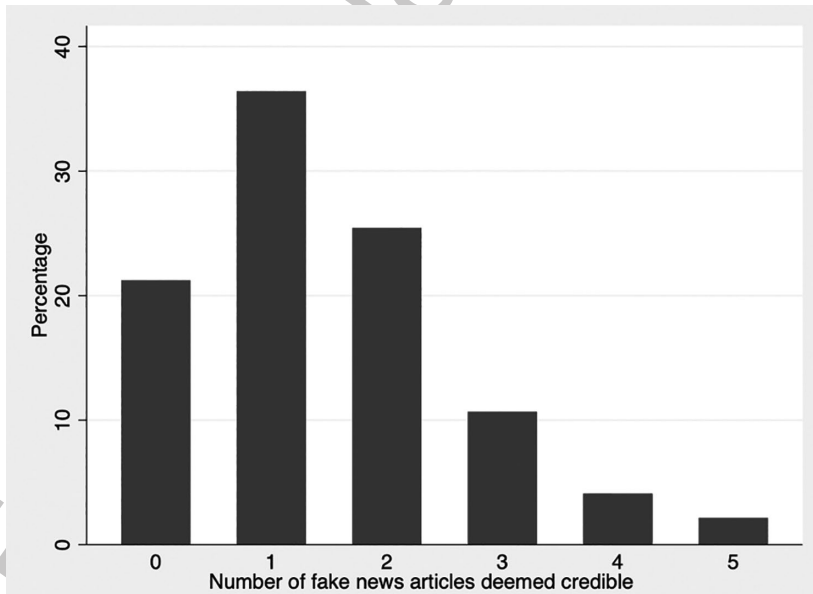


Figure 8.2 Proportion of fake news articles people find credible.

2.2 Calculation and Results

This section presents the results of a regression analysis to test whether endorsement of conspiracy theories or fake news is associated with intellectual vice. A regression approach goes beyond showing mere correlations between intellectual vice and the endorsement of questionable beliefs. Such correlations could be caused by some underlying third factor. Other explanations that have been suggested in the literature appeal to education, socio-economic background, political orientation, religion, and news consumption (Allcott and Gentzkow 2017; Brotherton et al. 2013; Hagen 2018; Lazer et al. 2018; Oliver and Wood 2014).

To vindicate intellectual vice, it should explain questionable beliefs over and above other, established explanations. Regression analysis allows us to test associations between outcomes and our measure of intellectual vice while controlling for these other factors.

Table 8.3 shows regression results for endorsement of conspiracy theories and fake news. Columns 1 and 2 concern conspiracy theories measured as the number of conspiracy theories endorsed. Columns 3 and 4 concern fake news endorsement measured as the number of fake news items deemed credible. We have normalized both outcome measures by calculating the z-score for each observation.

Columns 1 and 3 show regression results using only control variables. Controls used are age, household income, sex, education, ethnicity, political affiliation, religion, and news consumption. Coefficients can be compared with one another because all discrete variables have been normalized by computing their z-scores.

The reason to show these results is that they give us a baseline for how much of the variance in endorsement of conspiracy theories and fake news is accounted for by control variables. These regressions account for 22% and 26% of variance as measured by R^2 , respectively. In other words, our control variables can explain 22% of the variance between respondents in endorsing conspiracy theories, and controls explain 26% of variance in endorsing fake news.

Political affiliation and religion explain the most variance: Republicans were more likely to endorse conspiracies and fake news, as were more religious people. News consumption also plays a role. Respondents reading printed newspapers were less likely to believe conspiracies and fake news, respondents getting their news from social media were more likely. Households with higher income were less likely to endorse conspiracy theories, but income played no significant role in endorsing fake news.

Other variables show no statistically significant associations with conspiracy or fake news. Age, education, and ethnicity are not significantly associated with outcomes.

Columns 2 and 4 show regression results for the same set of controls plus the summary measure of intellectual virtue. For both outcomes,

Table 8.3 Regression results study 1

<i>Variables</i>	(1)	(2)	(3)	(4)
	<i>Conspiracy</i>	<i>Conspiracy</i>	<i>Fake News</i>	<i>Fake News</i>
Intellectual Virtue		-0.282*** (0.0323)		-0.245*** (0.0318)
Age	-0.0626* (0.0321)	-0.0380 (0.0310)	0.000221 (0.0322)	0.0216 (0.0309)
Income	-0.0997*** (0.0300)	-0.0885*** (0.0288)	-0.0313 (0.0277)	-0.0216 (0.0269)
Female	-0.0212 (0.0606)	0.0221 (0.0578)	0.114* (0.0589)	0.152*** (0.0570)
<i>Education</i>				
High school diploma or equivalent	-0.519** (0.249)	-0.612*** (0.209)	0.0308 (0.273)	-0.0497 (0.193)
Some college but no degree	-0.556** (0.242)	-0.662*** (0.199)	-0.0959 (0.265)	-0.189 (0.181)
Associate's degree	-0.617** (0.248)	-0.715*** (0.205)	-0.00911 (0.268)	-0.0947 (0.184)
Bachelor's degree	-0.564** (0.242)	-0.713*** (0.199)	-0.146 (0.264)	-0.275 (0.179)
Graduate degree	-0.564** (0.254)	-0.771*** (0.210)	-0.0641 (0.273)	-0.244 (0.189)
<i>Ethnicity</i>				
Asian or Pacific Islander	-0.320 (0.326)	-0.225 (0.338)	-0.232 (0.290)	-0.150 (0.246)
Black or African American	-0.297 (0.306)	-0.155 (0.324)	-0.130 (0.273)	-0.00607 (0.228)

(Continued)

	(1)	(2)	(3)	(4)
<i>Variables</i>	<i>Conspiracy</i>	<i>Conspiracy</i>	<i>Fake News</i>	<i>Fake News</i>
Hispanic	-0.475 (0.323)	-0.299 (0.337)	-0.298 (0.285)	-0.145 (0.242)
White/Caucasian	-0.409 (0.294)	-0.313 (0.311)	-0.0982 (0.260)	-0.0146 (0.214)
Other	0.0680 (0.376)	0.109 (0.377)	0.00276 (0.314)	0.0383 (0.267)
Political affiliation	0.272*** (0.0342)	0.226*** (0.0324)	0.341*** (0.0342)	0.302*** (0.0325)
Religion	0.222*** (0.0335)	0.219*** (0.0325)	0.199*** (0.0339)	0.196*** (0.0328)
<i>News Consumption</i>				
Newspapers	-0.0841*** (0.0297)	-0.0698** (0.0283)	-0.118*** (0.0315)	-0.106*** (0.0307)
Social Networks	0.134*** (0.0312)	0.0996*** (0.0299)	0.0746** (0.0304)	0.0449 (0.0292)
TV and Radio	-0.00735 (0.0303)	-0.0170 (0.0300)	0.0322 (0.0309)	0.0238 (0.0304)
Online Newspapers	0.0151 (0.0328)	0.0164 (0.0312)	0.00164 (0.0320)	0.00276 (0.0312)
News Aggregators	0.0181 (0.0301)	0.00686 (0.0296)	0.0692** (0.0299)	0.0594** (0.0297)
Constant	0.958** (0.379)	0.968*** (0.365)	0.139 (0.368)	0.148 (0.275)
Observations	949	949	949	949
R-squared	0.216	0.289	0.256	0.310

Robust standard errors in parentheses.
 *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

the association with intellectual virtue is statistically significant at the 1% level. The coefficient of intellectual virtue is larger than the coefficient of any of the controls. The proportion of variance we can explain jumps by 7 and 5 percentage points, respectively. This result suggests that intellectual vice explains endorsement of conspiracy theories and fake news over and above alternative explanations as measured by controls.

When intellectual virtue is added to the model, political affiliation and religion remain statistically significant, with coefficients at the same order of magnitude as intellectual virtue. It is noteworthy that sex becomes statistically significant at the 1% level for fake news endorsement once we account for intellectual virtue. Female respondents were somewhat more likely to endorse fake news.

2.3 Discussion

This exploratory study found that several demographic controls predict endorsement of conspiracy theories and credence in fake news. In addition, we found that intellectual virtue (and vice) account for a sizeable proportion of the variance in people's acceptance of both conspiracy theories and fake news. These results are correlational, so we cannot say for sure whether intellectual vice leads people to conspiracy theorizing or, instead, acceptance of conspiracy theories and fake news makes people intellectually vicious (or some third variable explains both). Nevertheless, the scale used to study intellectual virtue is meant to measure a trait, so the more plausible interpretation is that intellectual virtue explains conspiracy theorizing rather than the other way around. A longitudinal study might be able to shed further light on this question. In the next section, we describe a follow-up, pre-registered, confirmatory study that we conducted to further examine the relationship between intellectual character, on the one hand, and conspiracy theories and fake news, on the other hand.

3 Study 2

3.1 Materials and Methods

Study 2 replicates study 1. Participants were recruited using Amazon Mechanical Turk. The eligibility criteria were living in the United States and being 18 or older. Respondents were paid \$1.50 for participation. 1,011 people participated, of which 998 passed an attention check. Participants who failed the attention check were excluded from the analysis. Participants answered the same demographic questions as in study 1. Participants had an average age of 40 (SD = 13). 45% of the sample

was female. 79% of the sample was White/Caucasian, 10% was African American/Black, 5% was Asian or Pacific Islander, and 5% was Hispanic; the remaining 1% selected other or did not disclose ethnicity. 60% had obtained a bachelor's or a higher degree. Mean household income was 59,000 USD per year. All these demographic characteristics are very similar to the sample in study 1. Similarly, reported political affiliation, religion, and news consumption followed reports in study 1 closely.

Participants answered the same demographic questions and the same modules on intellectual virtue, conspiracist thinking, and fake news as in study 1. Average scores on the intellectual humility scale closely followed the scores in study 1, as did average numbers of fake news articles and conspiracy theories deemed credible. Table 8.9 in the appendix contains full descriptive statistics.

3.2 *Calculation and Results*

Table 8.4 shows regression results for conspiracy theories and fake news endorsement. Columns 1 and 2 replicate the analysis conducted in study 1 concerning conspiracy theories; columns 4 and 5 replicate the analysis conducted in study 1 concerning fake news endorsement. The results confirm the findings of study 1.

Columns 1 and 4 show regression results using only control variables. Controls are identical to study 1. All discrete variables have been normalized by computing their z-scores. These regressions account for 26% and 21% of variance as measured by R^2 , respectively. Control variables show broadly similar coefficients, with some noteworthy differences. Age was not a significant variable in the previous study. In this study, older participants are less likely to endorse conspiracy theories, significant at the 1% level. We note that this is in tension with the results of Guess et al. (2019), who found that older people disproportionately spread fake news. It might be that they spread it even though they don't believe it, but this tension calls for further investigation. Income is more consistently significant across regressions, at the 1% level: the richer you are, the less likely you are to endorse conspiracy theories and deem fake news credible. Gender is not significant in any of the regressions. Education remains mostly insignificant. Similarly to the original study, ethnicity is largely insignificant – with the exception of one ethnic group (American Indian or Alaskan Native), whose scores are based on just seven responses. Political affiliation and religion remain significant, with broadly similar coefficients to the original study. As in the original study, people who get their news from social networks were somewhat more likely to endorse conspiracy theories. The negative relationship between newspaper consumption and deeming fake news credible was replicated in this study.

Table 8.4 Regression results study 2

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Conspiracy	Conspiracy	Conspiracy	Fake News	Fake News	Fake News
Intellectual Virtue		-0.368*** (0.0347)			-0.320*** (0.0378)	
Open-mindedness			-0.328*** (0.0429)			-0.214*** (0.0444)
Modesty			-0.0557* (0.0299)			0.00866 (0.0345)
Corrigibility			0.131*** (0.0346)			0.0403 (0.0394)
Engagement			-0.210*** (0.0405)			-0.218*** (0.0413)
Age	-0.166*** (0.0320)	-0.0991*** (0.0289)	-0.107*** (0.0279)	-0.0855*** (0.0317)	-0.0271 (0.0296)	-0.0399 (0.0297)
Income	-0.169*** (0.0283)	-0.142*** (0.0263)	-0.130*** (0.0254)	-0.113*** (0.0309)	-0.0894*** (0.0284)	-0.0779*** (0.0283)
Female	-0.0943 (0.0584)	0.0173 (0.0545)	0.0557 (0.0517)	-0.0486 (0.0601)	0.0485 (0.0582)	0.0516 (0.0573)
<i>Education</i>						
High school diploma or equivalent	-0.224 (0.337)	-0.0871 (0.262)	-0.241 (0.285)	0.662** (0.303)	0.781** (0.317)	0.680** (0.280)
Some college but no degree	-0.226 (0.330)	-0.137 (0.253)	-0.209 (0.276)	0.555** (0.296)	0.632** (0.312)	0.605** (0.274)
Associate's degree	0.0778 (0.338)	0.108 (0.261)	-0.0217 (0.283)	0.681** (0.301)	0.707** (0.315)	0.626** (0.278)
Bachelor's degree	0.00514 (0.330)	-0.0447 (0.252)	-0.161 (0.275)	0.777*** (0.294)	0.734** (0.308)	0.677** (0.270)
Graduate degree	-0.0337 (0.335)	-0.104 (0.256)	-0.157 (0.279)	0.839*** (0.305)	0.778** (0.316)	0.775** (0.280)
<i>Ethnicity</i>						
American Indian or Alaskan Naive	1.726***	1.283***	1.010***	1.696***	1.311***	1.122***

(Continued)

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	Conspiracy	Conspiracy	Conspiracy	Fake News	Fake News	Fake News
Asian or Pacific Islander	(0.166) 0.00239	(0.136) -0.0214	(0.130) -0.0579	(0.218) -0.206	(0.187) -0.227**	(0.188) -0.248**
Black or African American	(0.137) 0.0955	(0.116) 0.177*	(0.121) 0.173*	(0.133) -0.0661	(0.114) 0.00438	(0.113) 0.0167
Hispanic	(0.101) 0.220	(0.0912) 0.273**	(0.0914) 0.264**	(0.109) 0.0640	(0.103) 0.110	(0.103) 0.108
Other	(0.136) -0.216	(0.119) -0.184	(0.119) -0.346	(0.148) -0.141	(0.128) -0.113	(0.130) -0.258
Political affiliation	(0.350) 0.190**	(0.353) 0.164**	(0.344) 0.168**	(0.318) 0.203**	(0.291) 0.181**	(0.315) 0.184**
Religion	(0.0301) 0.238***	(0.0275) 0.186**	(0.0265) 0.142**	(0.0324) 0.231**	(0.0312) 0.187**	(0.0302) 0.153**
	(0.0346)	(0.0313)	(0.0304)	(0.0373)	(0.0354)	(0.0349)
<i>News Consumption</i>						
Newspapers	-0.00277 (0.0307)	0.00546 (0.0292)	0.00319 (0.0284)	0.000930 (0.0325)	0.00809 (0.0312)	0.00959 (0.0304)
Social Networks	0.151*** (0.0316)	0.111*** (0.0302)	0.0994*** (0.0285)	0.0734** (0.0318)	0.0391 (0.0307)	0.0274 (0.0298)
TV and Radio	0.0225 (0.0302)	-0.0173 (0.0286)	-0.0135 (0.0283)	0.0908*** (0.0322)	0.0562* (0.0313)	0.0619** (0.0309)
Online Newspapers	0.0363 (0.0320)	0.0213 (0.0301)	0.0279 (0.0293)	0.0359 (0.0312)	0.0228 (0.0298)	0.0365 (0.0297)
News Aggregators	-0.0417 (0.0289)	-0.0462* (0.0275)	-0.0391 (0.0265)	0.0181 (0.0303)	0.0142 (0.0293)	0.0215 (0.0287)
Constant	0.0656 (0.329)	0.00807 (0.251)	0.100 (0.275)	-0.694** (0.295)	-0.744** (0.309)	-0.693** (0.271)
Observations	976	976	976	976	976	976
R-squared	0.264	0.375	0.427	0.207	0.291	0.323

Robust standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8.5 Regression results for individual virtues in study 2

	(1)	(2)	(3)	(4)	(5)
Variables	Conspiracy	Conspiracy	Conspiracy	Conspiracy	Conspiracy
Open-mindedness	-0.415*** (0.0345)				-0.328*** (0.0429)
Modesty		-0.142*** (0.0311)			-0.0557* (0.0299)
Corrigibility			-0.176*** (0.0307)		0.131*** (0.0346)
Engagement				-0.369*** (0.0321)	-0.210*** (0.0405)
Constant	1.251*** (0.263)	1.700*** (0.313)	1.658*** (0.271)	1.234*** (0.238)	1.111*** (0.262)
Observations	976	976	976	976	976
R-squared	0.405	0.280	0.292	0.380	0.427

Controls used but coefficients omitted: age, income, gender, education, ethnicity, ideology, religion, news consumption.

Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Columns 2 and 5 show regression results for the same set of controls plus the summary measure of intellectual virtue. As in the original study, the association with intellectual virtue is statistically significant at the 1% level for both conspiracy and fake news endorsement. The coefficients are larger than in the original study and larger than the coefficient of any of the controls. The proportion of variance we can explain jumps by 11 and 9 percentage points, respectively – significantly more than in study 1. This result confirms that intellectual vice explains endorsement of conspiracy theories and fake news over and above alternative explanations as measured by controls.

Columns 3 and 6 go beyond the analysis conducted in study 1. They show the coefficients of the four individual virtues that are aggregated in a single measure in columns 2 and 5. For both outcome variables, open-mindedness and engagement drive results. Intellectual modesty is insignificant concerning fake news endorsement. Corrigibility is insignificant concerning fake news. But it is noteworthy that corrigibility is significant at the 1% level concerning conspiracy theories – with a positive coefficient. More engaged participants were *more* likely to endorse conspiracy theories, other things equal. This result suggests that aggregating all four dimensions of the intellectual virtue scale is not advisable, since the dimensions may pull in opposite directions.

Let’s investigate the role of individual virtues further. We need to go beyond the multiple regression above because intellectual virtues are

highly correlated with one another (correlation coefficients between virtues range between 0.27 and 0.58). To see the effect of individual virtues, we therefore need to study their effect in isolation from other virtues. Column 5 in Table 8.5 shows the regression in column 3 above (coefficients of control variables are omitted for better readability).

Columns 1 to 4 analyse the association of individual virtues and the propensity to endorse conspiracy theories. Consistent with the results discussed above, open-mindedness and engagement are strongly negatively associated with conspiracy theorizing. But note that intellectual modesty and corrigibility also are, considered in isolation from the other virtues, negatively associated with conspiracy as well, at the 1% significance level, even if their coefficients are smaller. Hence we should read the positive sign of the engagement coefficient in column (5) as an artefact of the high correlation of engagement with the other virtues.

3.2.1 *Reexamining the Dimensions of Intellectual Humility*

The differentiated roles played by the four dimensions of the intellectual virtue scale in study 2 led us to reexamine the data from study 1. Do we find the same pattern in this independent sample? Note that this analysis must be considered exploratory because the data were collected before we arrived at our hypothesis. Nevertheless, the sample is independent and was not analyzed in this way until after study 2 had been analyzed, so it does shed some light on our topic.

Table 8.6 shows the results of the same regression as Table 8.5, applied to the data from study 1. The results are qualitatively similar: open-mindedness remains negatively associated with conspiracy theorizing, and the sign of corrigibility turns positive in the combined regression in column 5. Both modesty and engagement are negatively associated with conspiracy theorizing in the combined model. Yet all four virtues are negatively associated with conspiracist thinking and fake news endorsement. The coefficients are lower in study 1 than in study 2.

3.2.2 *Comparison Between Respondents with High and Low Intellectual Humility*

Table 8.7 compares respondents in the upper and lower half on the aggregate intellectual humility scale. The analysis is based on the combined datasets of studies 1 and 2 ($n = 1,973$). Consistent with the regression analysis, respondents in the high group engage less in conspiracist thinking and lend less credence to fake news. The difference is higher for conspiracist thinking than for identifying fake news, but both are significant at the 1% level.

Differences between a number of controls are also significant at the 1% level. Respondents scoring higher on intellectual virtue tend

Table 8.6 Regression results individual virtues study 1

	(1)	(2)	(3)	(4)	(5)
Variables	Conspiracy	Conspiracy	Conspiracy	Conspiracy	Conspiracy
Open-mindedness	-0.315*** (0.0324)				-0.270*** (0.0422)
Modesty		-0.120*** (0.0304)			-0.0973*** (0.0299)
Corrigibility			-0.0987*** (0.0312)		0.189*** (0.0352)
Engagement				-0.289*** (0.0327)	-0.210*** (0.0423)
Constant	1.030*** (0.354)	0.934** (0.384)	0.942*** (0.358)	1.031** (0.415)	1.083*** (0.414)
Observations	949	949	949	949	949
R-squared	0.309	0.230	0.225	0.294	0.345

Controls used but coefficients omitted: age, income, gender, education, ethnicity, ideology, religion, news consumption.

Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8.7 Demographic comparison combined sample

	Intellectual Virtue			
	Low	High	Difference	p
Fake news	2.74	2.44	0.29	0.00
Conspiracy	2.27	1.77	0.50	0.00
Age	38.49	41.01	-2.52	0.00
Income	58009	58776	-766	0.62
Female	0.43	0.54	-0.11	0.00
<i>Education</i>				
Less than a high school diploma	0.01	0.00	0.00	0.39
High school diploma or equivalent	0.08	0.12	-0.03	0.01
Some college but no degree	0.18	0.24	-0.06	0.00
Associate's degree	0.11	0.14	-0.03	0.02
Bachelor's degree	0.46	0.37	0.09	0.00
Graduate degree	0.17	0.13	0.04	0.01
<i>Ethnicity</i>				
American Indian or Alaskan Native	0.01	0.00	0.00	0.03
Asian or Pacific Islander	0.06	0.04	0.02	0.02
Black or African American	0.08	0.10	-0.02	0.21

(Continued)

	<i>Intellectual Virtue</i>			
	<i>Low</i>	<i>High</i>	<i>Difference</i>	<i>p</i>
Hispanic	0.05	0.05	0.00	0.69
White/Caucasian	0.78	0.80	-0.01	0.50
Other	0.01	0.01	0.00	0.72
Religion	2.55	2.36	0.19	0.00
Political Affiliation	3.53	3.19	0.33	0.00
<i>News Consumption</i>				
Printed Newspapers	3.05	3.11	-0.06	0.24
Social Networks	3.22	2.98	0.24	0.00
TV and Radio	3.39	3.34	0.05	0.35
Online Newspapers	3.03	3.01	0.02	0.70
News Aggregators	3.35	3.28	0.07	0.22

to be older. The age difference is 2.5 years on average. Moreover, the high group has a larger proportion of women than men, with a difference of 11 percentage points. Respondents in the high group are more likely to be Democrats, less likely to be religious, and less likely to get their news via social media. The differences between the groups are small for the education-related variables, and only some of them are significant. However, generally people with less formal education are likely to score higher on intellectual virtue. For none of the ethnic groups are differences between the high and low groups significant at the 1% level.

3.3 Discussion

In this study, we replicated the main finding of study 1: intellectual vice is strongly predictive, over and above many demographic controls, of acceptance of both conspiracy theories and fake news. After controlling for other variables, intellectual vice accounts for 10–13% of the variance in conspiracism and acceptance of fake news. In addition, we found that two of the four dimensions of intellectual character (open-mindedness and engagement) account for most of this effect.

As before, we cannot definitively conclude that intellectual character causes acceptance/rejection of conspiracy theories and fake news. It could be that people who get sucked in by conspiracy theories and fake news tend to become intellectually vicious, or that some third variable explains the correlations observed here. That said, it is at least plausible that the causal arrows run from traits (open-mindedness and engagement) to behaviors (accepting or rejecting conspiracy theories and fake news).

4 General Discussion

What can vice epistemologists learn from these results? The descriptive statistics show that questionable beliefs in fake news and conspiracy theories are widespread. Some established explanations account for some of the variances in outcomes. Republicans are more likely to endorse both conspiracy theories and fake news. Religion is also associated with conspiracist thinking and endorsement of fake news.

Intellectual vice appears to be an additional ingredient to the explanation. Intellectually vicious people are more likely to endorse conspiracy theories. This result holds up when we control for political orientation, religion, and a range of other factors. This finding supports claims by vice epistemologists that conspiracy theorists suffer from intellectual vice (Cassam 2016). Furthermore, intellectual vice is associated with questionable beliefs other than conspiracy theories. These experiments show that acceptance of fake news is associated with epistemic vice as well.

As we mentioned above, we think that the most plausible causal explanation of our findings is that intellectual character causes acceptance/rejection of conspiracy theories and fake news. However, we cannot rule out the opposite direction. Indeed, it may be that there are feedback loops at work here: intellectual vice makes people susceptible to conspiracy theorizing, which undermines their intellectual character still further, which makes them even more susceptible to conspiracy theorizing. Many readers will no doubt have an uncle or other family member who comes to mind in this context. Longitudinal studies could help to untangle these potential feedback loops.

In addition, we note that many, perhaps most, people encounter conspiracy theories and fake news primarily online. This study suggests that intellectual virtue and vice influence epistemic conduct in an online environment. Epistemic virtue appears to influence whether people place trust intelligently online (O'Neill 2002). As a next step, it would be interesting to compare the influence of intellectual virtue and vice in an offline setting.

We administered the scale as a self-assessment questionnaire. Self-assessment has two advantages vis-à-vis other methods of data collection in psychology and experimental philosophy. First, data gathering is relatively unproblematic. Through online services like Amazon Mechanical Turk, researchers have easy access to a large pool of participants (Buhrmester et al. 2011; Paolacci and Chandler 2014). Second, participants retain a high degree of autonomy over how they are described and rated. But this latter feature also gives rise to a challenge to self-assessment: the question of self-knowledge. Can we know our own vices? Ignorance about one's vices is a challenge because people can only set out to overcome their epistemic flaws once they recognize them. The self-assessment methodology is premised on the assumption that people have at least some

insight into their own character traits (Vazire 2010). The method does not require that people have a sophisticated conceptual understanding of intellectual virtue and vice. Rather, each intellectual trait is measured by aggregating responses to a number of agree-disagree items related to concrete behaviours, attitudes, motivations, and skills. Still, respondents may lack the self-knowledge necessary to respond to items adequately or may make up their minds on the spot (Dunning et al. 2004).

The philosophical correlate of the methodological problem with self-assessment is the problem of stealthy virtues and vices (Cassam 2015). Traits are stealthy if possessing the trait stands in the way of knowing that you have the trait. Self-knowledge about intellectual humility, in particular, may be tenuous. One feature of the truly humble may be that they do not think about themselves as particularly humble. The boastful, on the other hand, are unlikely to fully appreciate their lack of intellectual humility.⁵ In effect, the pretentious as well as the self-deprecatory may well lack the self-knowledge necessary to answer questions about their own intellectual humility correctly. Since intellectual virtue is however associated with epistemic outcomes as expected, people appear to have some knowledge about their own intellectual character traits. This finding is consistent with one of the validation studies for the survey used in this experiment. Alfano et al. conducted a study comparing self-ratings with ratings by informants and found positive correlations, though they managed to collect scores from only 107 informants (Alfano et al. 2017, 12ff.). This result suggests that the scale picks up on some traits of subjects that they and informers judge similarly.

Will participants respond to the items in the Intellectual Virtue Scale truthfully, even if they have self-knowledge? Participants who want to appear intellectually virtuous can easily do so by selecting socially desirable items, but in their validation study, Alfano et al. found that correlations between scores on the intellectual humility scale measures of socially desirable responding tended to be low (Alfano et al. 2017, 19f.). Perhaps this is because, in some communities, intellectual humility is not actually considered desirable. The transparency of the scale may limit its application to cases where respondents do not have strong incentives to answer in socially desirable ways. But in the absence of strong incentives to appear virtuous, the motive of self-discovery gives respondents a reason to answer truthfully. Since the scale relates to outcomes as predicted, the challenge of deception appears to be limited.

Finally, these results indicate that tracing the etiology of intellectual virtue and vice would be a valuable undertaking. How are open-mindedness, engagement, and the other intellectual virtues acquired? How are their corresponding vices acquired? At what ages do they first appear, and what causal factors influence their development? While methodologically challenging, developmental research into these questions would be highly revealing.

5 Conclusion

The results from the survey experiments suggest that intellectual virtue and vice are associated with the acceptance of fake news and conspiracist thinking. Intellectual virtue explains variance in the endorsement of conspiracy theories and fake news among respondents. Moreover, it has explanatory power over and above established explanations appealing to religiosity and political orientation.

The experiment makes a methodological contribution by showing that intellectual virtue and vice can be measured by a self-assessment scale. This result is supported by the finding that the self-reported measures of intellectual virtue and vice are related to epistemic outcomes in expected ways. People appear to have a good sense of whether they manifest the behaviours, attitudes, and motivations that reflect intellectual virtues or vices. Survey methodology can transform this knowledge into insights about intellectual vices.

The experiment demonstrates that empirical research can contribute to vice epistemology. Much remains to be done. Vice epistemologists have discussed intellectual vices including gullibility, dogmatism, prejudice, closed-mindedness, negligence, intellectual pride, idleness, cowardice, conformity, and rigidity. One important task is to develop a taxonomy of intellectual virtues and vices, and the social settings in which they are most important. Psychometric techniques provide compelling methods for developing such a taxonomy. We have only just begun to develop scales measuring individual intellectual virtues and vices. Eventually, experiments can contribute to answering the question of which intellectual vices matter most in which contexts. We need an empirical approach to investigate which intellectual vices are most harmful to gaining and transmitting knowledge and understanding.

Notes

- 1 For a skeptical treatment of the phrase and concept of ‘fake news’, see Coady (2019) and de Ridder (2019) responds to Coady’s skepticism.
- 2 To examine the registration, visit <https://osf.io/zbjgw>.
- 3 For an argument that conspiracy theorists are not irrational by their own standards, see Levy (2019). However, even if Levy is right, that does not mean that they are rational by objective, external standards.
- 4 For a full list of items, please refer to the validation paper by Alfano et al. (2017).
- 5 See, among others, Alfano and Robinson (2014) and Robinson and Alfano (2016).

References

- Alfano, M., and B. Robinson. 2014. “Bragging.” *Thought* 3(4): 263–272.
- Alfano, Mark, Kathryn Iurino, Paul Stey, Brian Robinson, Markus Christen, Feng Yu, and Daniel Lapsley. 2017. “Development and Validation of

- a Multi-Dimensional Measure of Intellectual Humility.” *PLoS One* 12(8): 1–28.
- Alfano, M., J. A. Carter, and M. Cheong. 2018. “Technological Seduction and Self-Radicalization.” *Journal of the American Philosophical Association* 4(3): 298–322.
- Alfano, Mark and Colin Klein. 2019. “Trust in a Social and Digital World.” *Social Epistemology Review and Reply Collective* 8(10): 1–8.
- Alfano, M., A. E. Fard, J. A. Carter, P. Clutton, and C. Klein. 2020. “Technologically Scaffolded Atypical Cognition: The Case of YouTube’s Recommender System.” *Synthese*.
- Allcott, Hunt, and Matthew Gentzkow. 2017. “Social Media and Fake News in the 2016 Election.” *Journal of Economic Perspectives* 31(2): 211–236.
- Brotherton, Robert, Christopher C. French, and Alan D. Pickering. 2013. “Measuring Belief in Conspiracy Theories: The Generic Conspiracist Beliefs Scale.” *Frontiers in Psychology* 4: 1–15.
- Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. “Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?” *Perspectives on Psychological Science* 6(1): 3–5.
- Cassam, Quassim. 2015. “Stealthy Vices.” *Social Epistemology Review and Reply Collective* (blog). October 16, 2015.
- . 2016. “Vice Epistemology.” *The Monist* 99(2): 159–180.
- . 2018. *Vices of the Mind*. Oxford University Press.
- . 2019. *Conspiracy Theories*. Polity.
- Coady, David. 2019. “The Trouble with ‘Fake News’.” *Social Epistemology Review and Reply Collective* 8(10): 40–52.
- de Ridder, Jeroen. 2019. “So What If ‘Fake News’ Is Fake News?” *Social Epistemology Review and Reply Collective* 8(10): 111–113.
- Dentith, Matthew R. X. 2016. “When Inferring to a Conspiracy Might Be the Best Explanation.” *Social Epistemology* 30(5–6): 572–591.
- Dunning, David, Chip Heath, and Jerry M. Suls. 2004. “Flawed Self-Assessment: Implications for Health, Education, and the Workplace.” *Psychological Science in the Public Interest* 5(3): 69–106.
- Gelfert, Axel. 2018. “Fake News: A Definition.” *Informal Logic* 38(1): 84–117.
- Guess, A., J. Nagler, and J. Tucker. 2019. “Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook.” *Science Advances* 5: eaau4586.
- Hagen, Kurtis. 2018. “Conspiracy Theories and the Paranoid Style: Do Conspiracy Theories Posit Implausibly Vast and Evil Conspiracies?” *Social Epistemology* 32(1): 24–40.
- Harman, Gilbert H. 1965. “The Inference to the Best Explanation.” *The Philosophical Review* 74(1): 88–95.
- Harris, Keith. 2018. “What’s Epistemically Wrong with Conspiracy Theorising?” *Royal Institute of Philosophy Supplements* 84: 235–257.
- Klein, C., P. Clutton, and V. Polito. 2018. “Topic Modeling Reveals Distinct Interests Within an Online Conspiracy Forum.” *Frontiers in Psychology* 9: 189.
- Klein, C., P. Clutton, and A. Dunn. 2019. “Pathways to Conspiracy: The Social and Linguistic Precursors of Involvement in Reddit’s Conspiracy Theory Forum.” *PLoS One* 14(11): 1–23.

- Lazer, David M. J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, et al. 2018. "The Science of Fake News." *Science* 359 (6380): 1094–1096.
- Levy, Neil. 2019. "Is Conspiracy Theorising Irrational?." *Social Epistemology Review and Reply Collective* 8(10): 65–76.
- Oliver, J. Eric, and Thomas J. Wood. 2014. "Conspiracy Theories and the Paranoid Style(s) of Mass Opinion." *American Journal of Political Science* 58(4): 952–966.
- O'Neill, Onora. 2002. *A Question of Trust: The BBC Reith Lectures 2002*. Cambridge University Press.
- Paolacci, Gabriele, and Jesse Chandler. 2014. "Inside the Turk: Understanding Mechanical Turk as a Participant Pool." *Current Directions in Psychological Science* 23(3): 184–188.
- Roberts, R., and J. Wood. 2007. *Intellectual Virtues*. Oxford University Press.
- Robinson, B., and M. Alfano. 2016. "I Know You Are But What Am I? Anti-Individualism about Intellectual Humility and Wu-Wei." *Logos & Episteme* 7(4): 435–459.
- Vazire, Simine. 2010. "Who Knows What about a Person? The Self–Other Knowledge Asymmetry (SOKA) Model." *Journal of Personality and Social Psychology* 98(2): 281.

Appendix

Table 8.8 Summary statistics study 1

<i>Variable</i>	<i>Obs</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Min</i>	<i>Max</i>
Intellectual Virtue	975	67.66	12.60	25.00	100.00
Open Mindedness	975	80.08	14.51	16.67	100.00
Modesty	975	52.96	19.81	0.00	100.00
Corrigibility	975	66.36	17.85	5.00	100.00
Engagement	975	68.55	16.71	8.33	100.00
Conspiracy Theories	975	1.96	0.95	1.00	5.00
Fake News	975	2.55	0.70	1.00	5.00
Age	974	39.74	12.73	24.00	74.00
Income	958	57474	34368	10000	125000
Female	968	0.52	0.50	0.00	1.00
<i>Education</i>					
High school diploma or equivalent	974	0.10	0.30	0.00	1.00
Some college but no degree	974	0.22	0.41	0.00	1.00
Associate's degree	974	0.15	0.35	0.00	1.00
Bachelor's degree	974	0.39	0.49	0.00	1.00
Graduate degree	974	0.14	0.35	0.00	1.00
<i>Ethnicity</i>					
Asian or Pacific Islander	971	0.05	0.22	0.00	1.00
Black or African American	971	0.09	0.28	0.00	1.00
Hispanic	971	0.05	0.22	0.00	1.00
White / Caucasian	971	0.79	0.41	0.00	1.00
Other	971	0.01	0.11	0.00	1.00
Political Affiliation	968	3.36	2.08	1.00	7.00
Religion	973	2.44	1.51	1.00	5.00
<i>News Consumption</i>					
Printed Newspapers	975	3.11	1.22	1.00	5.00
Social Networks	975	3.13	1.28	1.00	5.00
TV and Radio	975	3.38	1.24	1.00	5.00
Online Newspapers	975	3.09	1.13	1.00	5.00
News Aggregators	975	3.34	1.20	1.00	5.00

Table 8.9 Summary statistics study 2

<i>Variable</i>	<i>Obs</i>	<i>Mean</i>	<i>Std. Dev.</i>	<i>Min</i>	<i>Max</i>
Intellectual Virtue	998	67.10	14.46	10.23	100.00
Open Mindedness	998	78.44	16.49	16.67	100.00
Modesty	998	53.59	20.69	0.00	100.00
Corrigibility	998	66.60	19.02	0.00	100.00
Engagement	998	67.44	18.20	8.33	100.00
Conspiracy Theories	998	2.09	1.03	1.00	5.00
Fake News	998	2.65	0.75	1.00	5.00
Age	998	39.64	13.02	24.00	74.00
Income	990	59245	34422	10000	125000
Female	994	0.45	0.50	0.00	1.00
<i>Education</i>					
High school diploma or equivalent	998	0.10	0.30	0.00	1.00
Some college but no degree	998	0.20	0.40	0.00	1.00
Associate's degree	998	0.10	0.31	0.00	1.00
Bachelor's degree	998	0.44	0.50	0.00	1.00
Graduate degree	998	0.16	0.36	0.00	1.00
<i>Ethnicity</i>					
Asian or Pacific Islander	997	0.05	0.21	0.00	1.00
Black or African American	997	0.10	0.30	0.00	1.00
Hispanic	997	0.05	0.21	0.00	1.00
White/Caucasian	997	0.79	0.41	0.00	1.00
Other	997	0.01	0.11	0.00	1.00
Political Affiliation	988	3.38	2.07	1.00	7.00
Religion	993	2.48	1.47	1.00	5.00
<i>News Consumption</i>					
Printed Newspapers	998	3.05	1.23	1.00	5.00
Social Networks	998	3.08	1.33	1.00	5.00
TV and Radio	998	3.35	1.19	1.00	5.00
Online Newspapers	998	2.95	1.13	1.00	5.00
News Aggregators	998	3.30	1.20	1.00	5.00

8b Commentary from Quassim Cassam

Reply to Marco Meyer and Mark Alfano

The hypothesis that intellectual vice predicts and explains the acceptance of conspiracy theories and fake news is not new. What is new and valuable in Meyer and Alfano's chapter is the attempt to put this hypothesis to the test. Unlike armchair vice epistemologists (myself included), Meyer and Alfano are not content to rely on intuition in positing a link between intellectual vice and conspiracy theorizing. They want empirical evidence, and they find it. For those of us who have always suspected that there is such a link, this is a heartening result. It is nice when something that one believed to be the case turns out, on further study, actually to be the case.

The main finding of the two studies carried out by Meyer and Alfano is that "intellectual vice is strongly predictive, over and above many demographic controls, of acceptance of both conspiracy theories and fake news" (p. 27). Political affiliation and religion are also relevant, and Meyer and Alfano are careful not to conclude definitively that intellectual character *causes* acceptance of conspiracy theories and fake news. However, the existence of a causal link is a plausible explanation of their findings. Their research also makes an important methodological contribution by showing that "intellectual virtue and vice can be measured by a self-assessment scale" (p. 31). The implication is that intellectual vices are not fully "stealthy", where a stealthy trait is defined as one that stands in the way of knowing that you have the trait.

While these findings are welcome, it would be worth reflecting on how they bear on the following issue: conspiracy theorists presumably do not believe *every* conspiracy theory, even if they believe more than one. Thus, one might ask: *which* conspiracy theories are they likely to believe? Uscinski and Parent suggest that "people's political ideologies play a strong role in determining which conspiracy theories they will subscribe to" (2014, 12). For example, they cite research which found that free-market ideologies predicted the endorsement of climate change conspiracy theories. This finding is unsurprising since "those believing in free markets would prefer not to endorse the collectivist policies that

are proposed for combatting climate change” (Uscinski and Parent 2014, 13). In much the same way, one would not be surprised to find that the Great Replacement conspiracy theory – the theory that white Europeans are being deliberately replaced through immigration – is popular among white supremacists, because this theory is integral to the ideology of white supremacism.

One conclusion that one might draw from this is that politics is more important than intellectual character when it comes to predicting belief in conspiracy theories. Meyer and Alfano show that such a conclusion would be too hasty: political affiliation and intellectual character *both* matter. Still, it is worth noting the ideological associations of the conspiracy theories they use to elicit the propensity to endorse conspiracy theories. Two of the five theories listed by Meyer and Alfano – that the invasion of Iraq was driven by Jews in America and Israel and that George Soros is behind a plot to destabilize the American government – are anti-Semitic. They operate in what Jovan Byford calls an “ideological space with a long antisemitic tradition” (2011, 100). The theory that President Obama was not born in America, which is another one of the examples used by Meyer and Alfano, is not anti-Semitic but it is arguably racist. It is interesting that race plays such a significant role in three of the five theories cited by Meyer and Alfano. Indeed, the theory that the financial crisis was engineered by wealthy bankers – number four on Meyer and Alfano’s list – also has racist overtones, to the extent that terms like “banker”, “financier”, and “globalist” are often used in the world of conspiracy theories as code for “Jew”.

This raises an important question about characterological approaches to conspiracy theorizing: are they in danger of downplaying the ideological drivers of belief in conspiracy theories? Specifically, are they in danger of downplaying the racist element in several of the most prominent modern conspiracy theories? Since Meyer and Alfano are careful not to ignore the role of political affiliation in conspiracy theorizing, they cannot be accused of making this error. Nevertheless, there is a question of emphasis. In focusing on the intellectual vices of conspiracy theorists, is there a danger of de-emphasizing the ideological drivers and political agendas of those who put forward toxic conspiracy theories? To the extent that acceptance of conspiracy theories is integral to extremist ideologies, it seems likely that the factors that draw people to conspiracy theories are closely related to the drivers of political extremism.

There is also another point that needs to be highlighted. People who accept conspiracy theories are predominantly conspiracy theory *consumers*. However, there are also conspiracy theory *producers*, people who invent and propagate conspiracy theories. Sunstein describes them as “*conspiracy entrepreneurs*” (2014, 12). There is no need to suppose that the person or persons who fabricated the *Protocols of the Elders of Zion* believed their own conspiracy theory. To ask why they believed the

Protocols is to ask the wrong question. The real question is: why would anybody want to invent and circulate such a tissue of lies? The obvious answer is political or ideological: the *Protocols* advanced their ideological agenda. In the same way, the Great Replacement advances the agenda of white supremacy, and conspiracy theories about school shootings in the United States are designed to deflect arguments for tighter gun control. A person's intellectual vices might help to explain their support for repellent political ideologies, but it is hard to avoid the conclusion that their main failing is moral rather than intellectual. They are not just bad thinkers but bad people.

References

- Byford, J. (2011), *Conspiracy Theories: A Critical Introduction* (Basingstoke: Palgrave Macmillan).
- Sunstein, C. (2014), *Conspiracy Theories and Other Dangerous Ideas* (New York: Simon & Schuster).
- Uscinski, J. & Parent, J. (2014), *American Conspiracy Theories* (Oxford: Oxford University Press).

8c Commentary from Colin Klein

The Virtues You Project and the Vices You Have: Commentary on Meyer and Alfano

Meyer and Alfano (this volume) present intriguing evidence of a correlation between intellectual vice and endorsement of both conspiracy theories and fake news. This is surprising and informative. As they note, this goes against the notion – argued by Cassam (2019) among others – that epistemic vices might be “stealthy”, and so unavailable to those who have them. The specific pattern of results is also interesting. Meyer and Alfano found that the effect was driven primarily by two vices: the opposite of Open-Mindedness (which they term, “Intellectual Arrogance”) and the opposite of Engagement (“Boredom”).

What might explain Meyer and Alfano’s results? They note two plausible causal stories. The first is that intellectual vice causes belief in conspiracy theories: bad reasoning leads to bad results. The second is that conspiracy theorizing leads to vice. Conspiracy theories claim that evidence about powerful behind-the-scenes machinations is being distorted by the very agents who work their nefarious plans. If you really believed this, why not feel superior for having noticed it? Why not feel a bit uninterested in finding out more? (It would all be lies anyway!)

There is a third option that I think is also worth considering. First, some background. Most of the current research on conspiracy theories has examined them as they play out in online forums (Wood and Douglas 2015). As someone who has spent an unreasonable amount of time reading comments on conspiracy forums, I can say that Meyer and Alfano’s particular pattern of results was striking. Online, you find a consistent theme: conspiracy theorists present themselves as truly open-minded and engaged, willing to question the mainstream by seeking new sources of information, and so on *ad nauseam*.

Granted, that is anecdotal – though as Klein et al. (2018) point out, conspiracy theorists do spend quite a bit of time explicitly talking about evidence and sources of evidence, which gives some support to that initial impression. I note that Meyer and Alfano also found a smaller but

significant *positive* correlation with modesty. Again, this is striking. Intellectual modesty is not typically something that one associates with conspiracy theorists; if anything, they seem more likely to reside down in the Dunning-Kruger tarpit.

The apparent discrepancy has a straightforward reconciliation. Committed online conspiracy theorists often take their beliefs to be part of an identity that connects them to a broader community of like-minded individuals (Franks et al. 2017). Identity formation in the context of a group involves a kind of *performance* for others (Goffman 1959), one which marks oneself as a member of the ingroup and delineates a role within it. Part of that role is presenting oneself as a heroic seeker of the truth – the sort of person who likens themselves to “Socrates, Jesus of Nazareth, Giordano Bruno, and Galileo Galilei... so dangerous that authorities tried to silence them” (McMahan et al. 2021).

Goffman points out that when someone plays a role, “he implicitly requests his observers to take seriously the impression that is fostered before them. They are asked to believe that the character they see actually possesses the attributes he appears to possess...” (1959, 17). Online forums are remarkably welcoming in this regard. Yet there are certain virtues that are very hard to actually develop by pretending that you have them – and indeed, espousing them can cut against their development. Humility is obviously like this: one doesn’t become humble by talking about one’s own humility all the time. Similarly, I suggest, this might be the case with many epistemic virtues. Talking about how close-minded *others* are is not a great way to develop one’s open-mindedness. Echo chambers are not a great place to develop good habits of engagement. So, what Meyer and Alfano show is that self-report in isolation is reasonably telling, but that it can come apart from the sort of epistemic grandstanding that characterizes online discussions.

The three causal stories I have sketched are not mutually exclusive. One might find the sort of vicious cycle familiar from other domains (e.g., depression leads to anxiety about social situations, which creates avoidant behavior, which leads to isolation, which is depressing). The possible causal loops might be more indirect still. At the level of the mainstream media, conspiracy theories are primarily promulgated by politically conservative outlets (Benkler et al. 2018). Insofar as epistemic vice is also correlated (presumably contingently) with conservatism and strong religious beliefs, and those are correlated (again presumably contingently but currently) with epistemic vice, there might be all sorts of higher-order feedback. The appropriate sort of causal story for epistemic vice, as with other sorts of vices, might thus be more of a complex epidemiological one (Eaton 2007). Part of that story, I suggest, may be how social dynamics drive a wedge between one’s self-presentation as epistemically virtuous and the actual development of epistemic virtues.

References

- Benkler, Y., Faris, R., and Roberts, H. (2018) *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press.
- Cassam, Q. (2019) *Vices of the Mind: From the Intellectual to the Political*. Oxford: Oxford University Press.
- Eaton, A.W. (2007) A Sensible Antiporn Feminism. *Ethics*. 117(4): 674–675.
- Franks, B., Bangerter, A., Bauer, M.W., Hall, M., and Noort, M.C. (2017) Beyond “Monologicality”? Exploring Conspiracist Worldviews. *Frontiers in Psychology*. 8: 861.
- Goffman, E. (1959) *The Presentation of Self in Everyday Life*. New York: Anchor Books.
- Klein, C., Clutton, P., and Polito, V. (2018) Topic Modeling Reveals Distinct Interests within an Online Conspiracy Forum. *Frontiers in Psychology*. 9: 189.
- McMahan, J., Minerva, F., and Singer, P. (2021) Editorial. *Journal of Controversial Ideas*. 1(1), 11.
- Wood, M.J., and Douglas, K.M. (2015) Online Communication as a Window to Conspiracist Worldviews. *Frontiers in Psychology*. 6: 836.

8d Marco Meyer and Mark Alfano's Response to Commentaries

Response to Comments By Colin Klein and Quassim Cassam on Fake News, Conspiracy Theorizing, and Intellectual Vice

We are grateful for the thoughtful and thought-provoking challenges by Colin Klein and Quassim Cassam. Klein suggests an intriguing distinction between actual and espoused epistemic virtue. He challenges us to think about what we are really measuring with the survey instrument we employed. Cassam pushes us to think more carefully about what conspiracy theories really are. He raises the question whether studying conspiracy theories in abstraction from their usually antisemitic and racist ideological background obscures something fundamentally important about them.

Let's turn to Klein's challenge first. He scrutinizes the correlations that we find between epistemic vice, especially intellectual arrogance and boredom, and belief in conspiracy theories. We considered two possible explanations in our article: most theorizing about epistemic vice would suggest that epistemic vice leads people to buy into conspiracy theories. But based on our correlation results we cannot rule out that the causal arrow runs instead in the opposite direction: conspiracy theorizing may make respondents more intellectually vicious.

Klein challenges us to complicate the picture by suggesting that epistemic virtue and vice may be performative. Given that in online forums, conspiracy theorists spend a lot of time explicitly scrutinizing evidence, how does that sit with our suggestions that they are bored and arrogant? He suggests that conspiracy theorists often appear to be committed to a community of the like-minded and that they tap into a conspiracist identity which requires presenting yourself as truly open-minded. Perhaps conspiracy theorists see themselves as conditionally open-minded and are only willing to engage in (what they consider to be) open-minded inquiry with other conspiracy theorists. In all of us, espoused and actual epistemic virtues and vices may come apart.

If Klein is right in suggesting that this gap between actual and espoused epistemic virtues and vices is particularly large for conspiracists,

two questions emerge: does our survey instrument measure the former or the latter? And which of the two, if any, is causally efficacious? We share Klein's observation that our findings that conspiracy theorists are bored and arrogant is out of sync with their self-presentations. Since findings seem to suggest that there is a relation between what the survey is measuring and conspiracy theorizing, this would suggest our survey is picking up on actual rather than espoused virtues and vices.

Klein suggests an interesting additional research question: Are espoused ethical virtues and vices associated with conspiracy theorizing also? We would not be surprised. One of us has argued that epistemic virtues and vices can have effects on behavior even if they are merely espoused (Alfano 2013). One way of testing this in future research would be to mine social media comments for displays of epistemic virtue and vice and take this as a measure of espoused epistemic virtues and vices.

Quassim Cassam puts another important challenge to us: does our methodology run the risk of de-emphasizing the ideological drivers and political agendas of conspiracy theorists? The challenge has two parts. First, Cassam urges us to consider not only the consumers, but also the producers of conspiracy theories. Even if epistemic vice can explain the gullibility of conspiracy theory consumers, is epistemic vice an equally good explanation for what drives producers? We agree with Cassam that we should not assume producers believe their conspiracist concoctions. Often the attempt to support racist or antisemitic political ideologies is a more likely explanation than epistemic vice – though a disposition to lead others to accept misinformation or to confuse them is surely an *other-regarding* epistemic vice. Our study is firmly focused on consumers of conspiracy theories. However, further research might investigate how sharp the dividing line between conspiracy consumers and producers really is. Anecdotally, people who primarily consume conspiracy theories may at the same time produce local or ad hoc conspiracy theories.

Second, considering the consumers of conspiracy theories, Cassam observes that popular conspiracy theories like the ones we test have important features in common. Notably, they are very often antisemitic and/or racist. Cassam's challenge is that our emphasis on epistemic virtues and vices may distract from other, more important moral and political drivers of political extremism. We acknowledge that there are other important drivers of conspiracy theorizing. We were interested in showing whether epistemic virtue and vice can explain the remaining differences once we account for political, demographic, and other psychological drivers of conspiracy theorizing. Still, it is a fair question to what extent our survey predicts bad habits of thought, or to what extent it predicts racist and antisemitic ideology. One way of investigating that question is to test whether epistemic virtue and vice are associated with invented conspiracy theories, which are not (explicitly) antisemitic or

racist – though we must admit that lurking behind many such conspiracy theories there is often still an element of antisemitism or racism.

Perhaps one way of characterizing the relationship between conspiracist thinking and epistemic virtue and vice is this: conspiracy theories are attempts to pollute our thinking, usually to tempt us to adopt antisemitic or racist ideologies or at least not to oppose them as vigorously and firmly as we otherwise would (e.g., Holocaust denialism). The attractiveness of this temptation is perhaps best explained by sociological categories. But epistemic virtue matters because it may function as a corrective, enabling people to resist conspiracist thinking. By contrast, epistemic vice may make people more prone to give in to conspiracist thinking. More research is needed to establish, however, that the mechanism that produces the correlations that we find is a causal mechanism running from epistemic vice to conspiracist ideation.

Reference

Alfano, M. (2013). *Character as Moral Fiction*. Cambridge University Press.

9 Playfulness Versus Epistemic Traps

C. Thi Nguyen

Dogmatism often seems to come packaged with a mood of grim and unpleasant humorlessness. And when dogmatists do indulge in humor, it's often of a decidedly heavy-hearted sort: smug mockery and harsh satire. At least in the popular imagination, dogmatism does not seem to sit easily with a spirit of genuine lightheartedness or play.

And we can find various playful qualities—lighthearted humor, a sense of fun—associated with a more intellectually fluid mode of being. Laughter and play may not be required for all forms of intellectual achievement, but they are strongly associated with some particular forms of intellectual virtue. The joking genius, the laughing sage—these are all familiar archetypes. Of course, these might just be stereotypes or cultural mythologies. But might these popular associations reveal some kind of genuine and deep connection between playfulness and intellectual virtue?

In this chapter, I'll take a reconstructive approach. Let's look to see if there might be some plausible cognitive function for playfulness, some way in which it might help us in our struggles to cope with and understand the world. But if one surveys the literature on intellectual virtue, the ideal which emerges is a figure who is, if not actively dour, then at least not very much fun. Here's a typical example, from a contemporary discussion of intellectual virtue: "the most excellent cognizer" turns out to be "sober, careful, conscientious, thorough, and the like" (Riggs 2010, 184). There are certainly people in whom intellectual virtue emerges in such a sober manner. But that description seems to leave out other approaches to being a thoughtful and sensate person. Some sages are full of humor, and some of the best insights start as jokes.

Here, I'll take the first step towards an account of one particular virtue in this space: the epistemic virtue of *intellectual playfulness*. Intellectual playfulness, loosely, is the disposition to try out new ideas, perspectives and systems of thought for the sheer joy of it. Intellectual playfulness, I will argue, is the right disposition to get us out of a certain kind of dogmatism. This isn't its only role in our lives. Playfulness is surely valuable in and of itself—a source of joy and laughter. But intellectual playfulness also has some clear epistemic functionality for us.

Intellectual playfulness, I will suggest, is a disposition to explore ideas for the value of the exploration itself. The ramblings of intellectual playfulness are not guided, in their particular movements, by a hope of finding a truer and better theory. The intellectually playful person tries out ideas because the process is fun or pleasingly wild, or because the ideas are beautiful. In this way, it is a distinctive process from the intellectual exploration of the truth-seeker. The truth-seeker's explorations are guided by the current belief system; they will typically check out the most plausible alternatives. The intellectually playful person doesn't care about plausibility. They care about more aesthetic qualities of ideas. They care about cool ideas, or elegant ones, or thrilling joy rides of discovery. They care about exploring where exploration is joyful.

I will suggest that the intellectually playful exploration sometimes can better serve the goal of finding the truth than will exploration that is strictly aimed at finding the truth. The best approach to finding out the truth will turn out to include some joyful rambles away from it. To bring out the value of intellectual playfulness, I will show how it functions against one of its natural enemies: epistemic traps. Epistemic traps are belief systems that undermine our epistemic efforts, leaving us stuck inside them. Intellectual playfulness is the right disposition to get us out of such a trap, if we happen to fall in one. And since it is hard to tell if one is in such a trap, it's good to maintain some intellectual playfulness at all times. Playfulness is an intellectual insurance policy.

1 Epistemic Traps

To understand the value of intellectual playfulness, then, we'll need to get a clearer view of how epistemic traps work. So: some belief systems linger because they are epistemically successful. They contain a starting seed of good beliefs and help us to find more good beliefs.¹ But other belief systems linger, not because they guide us toward the truth, but because they are sticky. I am particularly interested in those belief systems that linger because they work to prevent their believers from seeing or acknowledging good contrary evidence. Such belief systems seem rigged up to block defection. Let's call these belief systems *epistemic traps*. (By *belief system* I don't just mean a set of beliefs about propositions, but also the values that guide the acquisition and evaluation of particular beliefs.)

Some trap belief systems operate by preventing their adopters from reflecting on their belief system at all. They prevent, in their adopters, processes like evidence-gathering, reflection, and deliberation.² We can call these *antireflective traps*. One example: a belief system that emphasizes unswerving and unthinking obedience to a leader—a deference trap.³ Another example: a belief system that made its followers so

undermotivated in general that they lost the verve to reflect at all—an apathy trap.⁴ Other examples include belief systems that encourage one to drug themselves into oblivion, or starve oneself, or exercise so vigorously so as to obliterate all thought.

Even more insidious than anti-reflective traps, however, are those belief systems that encourage, but *redirect*, various intellectual processes—leading good-faith, epistemically oriented efforts astray. Such a belief system performs a kind of intellectual judo, flipping earnest intellectual efforts and sending down the wrong paths. They are traps for active inquiry.

Let's call something an *inquiry trap* if it has the following characteristics:

- 1 It is a belief system (including some set of beliefs and relevant norms, values, and standards for evaluating, adopting, and discarding beliefs).
- 2 It is arranged such that good-faith, epistemically-oriented attempts at inquiry are redirected to yield epistemically poor results.
- 3 Those poor results tend to reinforce the belief system.

Anti-reflective traps discourage the process of wriggling to find the truth for yourself. Inquiry traps redirect that wriggling, pulling you more tightly into the trap.⁵

One example of an inquiry trap is the belief system associated with an *echo chamber*.⁶ Echo chambers are social structures that bring insiders to distrust all outsiders. I have discussed echo chambers at length elsewhere. To summarize: an echo chamber is a community that creates a significant trust disparity between members and nonmembers. That disparity is created by undermining the credibility of nonmembers and amplifying the credibility of members. Echo chambers also come with a core belief system, which one must accept to count as a member. Crucially, that belief system includes beliefs that maintain and increase that trust disparity.

I draw my analysis from Kathleen Hall Jamieson and Frank Cappella's landmark empirical study of the right-wing echo chamber around Rush Limbaugh and Fox News. Limbaugh's followers adopted the belief-system promulgated by Limbaugh. That belief system includes the view that everybody who didn't share those views was caught in the grips of a corrupt media, which had been taken over by malicious liberal elites (Jamieson and Cappella 2010).⁷ Though Jamieson and Cappella's analysis is of a right-wing political echo chamber, we can find examples of echo chambers among liberals and among centrists, and across all manner of nonpolitical domains. I believe I've seen echo chambers around particular forms of exercise, breastfeeding theories, systems of nutrition, and science denialism.

It's crucial that we distinguish echo chambers from a nearby phenomenon: that of epistemic bubbles. An epistemic bubble is a social structure that *omits* outsider voices, while an echo chamber is a social structure that *discredits* outsider voices. Epistemic bubbles leave their insiders ignorant of relevant evidence; echo chambers leave their members actively distrustful of outside sources. Current usage often conflates these two ideas—usually ignoring the possibility of trust manipulation, and focusing on epistemic-bubble-style filtration effects. But epistemic bubbles aren't the most significant threat right now. Epistemic bubbles pop easily; we simply need to expose insiders to the evidence that they've missed. Echo chambers are much more robust. Members of echo chambers come equipped with the intellectual machinery needed to dismiss contrary evidence coming in from the outside. Outside sources are, after all, untrustworthy, malicious, and corrupt.⁸

Notice that epistemic bubbles aren't inquiry traps, but echo chambers are paradigmatic inquiry traps. Epistemic bubbles do entrap their members, but they work through bad connectivity in their external information delivery network. An echo chamber, on the other hand, changes how inquiry will go by discrediting outside sources. A member's attempts to understand the truth will immediately run afoul of the echo chamber's trust settings, which will guide them to dismiss many reliable informants and trust many unreliable informants. But notice that the echo chamber member isn't unreflective or unthinking. They are often furiously analyzing incoming information—seeing where it comes from, and deploying their background theories about who's trustworthy and who's malicious. Echo chambers can furnish their members with vigorous and satisfying intellectual lives, since the belief system makes it easy for them to create powerful, seemingly-apt and seemingly-unifying explanations for all manner of phenomena.⁹

Echo chambers also typically contain *disagreement-reinforcement mechanisms*. For example, the leader of an echo chamber might claim that everybody on the outside was part of some vast conspiracy to undermine our country—and that those conspirators will try to corrupt the true believers by undermining the leader, with fake contrary evidence, or stories about the leader's corruption and unreliability.¹⁰ Often, these mechanisms involve conspiracy theories that implicate journalists, universities, scientists, or other external sources of information.¹¹ Thus, echo-chamber members are prepared for assaults from the outside, with pre-established machinery designed to dismiss contrary evidence from the outside. Endre Begby calls this process *evidential pre-emption* (Begby 2020). Crucially, Begby points out, evidential pre-emption not only disarms incoming evidence, but can create a positive feedback loop inside the echo chamber. The leader has made a prediction: that outsiders will try to undermine the leader's authority. When outsiders do try to undermine that leader, then, from the perspective of the insiders,

the leader's predictions have come true—which is a reason to increase their trust in their leader. Disagreement-reinforcement mechanisms are a truly elegant piece of malicious design. With such a mechanism, an echo chamber's defenses also serve to simultaneously increase the echo chamber's grip.

It is tempting to attribute to our political opponents' pure unthinkingness or brute idiocy. But I think that inquiry traps are far more common than brute unthinkingness. Pure unthinkingness is easier to detect and to recognize as problematic. Inquiry traps are more insidious precisely because they permit—and often foster—vigorous intellectual effort. They help create, in their members' self-inspection, the appearance of intellectual virtue.

Some epistemic traps hybridize the strategies of antireflective traps and inquiry traps. Consider what we might call an *insensitivity trap*. An insensitivity trap is a belief system that selectively cuts off attention to certain areas of life by attributing valuelessness to those areas. This typically occurs by narrowly specifying what counts as valuable. Consider, for example, the archetypical figure of the businessperson who believes that the only thing of any importance is money. Since they care only about money, they are unlikely to notice many of the things that might have pressured them to revise their belief system. They are likely to spend all their time thinking about strategies to make more money, and unlikely to spend any time on, say, literature or various humanistic pursuits. They fail to attend to the very pursuits which might put them into contact with other expressions of value.¹² Similarly, imagine a philosopher who thinks that the only worthwhile philosophy is well-articulated and rigorously developed, and which addresses a carefully delineated set of topics. Such a philosopher will ignore anything that lacks that style of articulation, or which addresses a different set of topics. They will fail to adequately attend to ideas and expressions that might have served to broaden their sense of what was worthwhile.

Notice that the insensitivity trap shares with the antireflective trap a certain stifling of key reflective processes. Our insensitive businessperson doesn't ask, say, philosophical questions about the value of a life spent with money, because their belief system has rendered such questions valueless. Our insensitive philosopher doesn't ask questions about, say, systematic oppression, since those questions cannot be well-articulated inside their designated set of worthy topics—so the topic appears uninteresting. But the insensitivity trap also shares with the inquiry trap a quality of redirection. Our businessperson could be spending plenty of time assessing their belief system and fine-tuning their beliefs, as they optimize their ability to make money. But those efforts are all spent in a narrowed and focused direction, as set by their belief system. The businessperson is not utterly unreflective; rather, their efforts of reflection have been channeled along sharply delineated paths. They might be

extremely reflective about, say, rooting out those cognitive biases which make them worse at investing, but entirely unreflective about why their life has nothing in it but financial pursuits.

To simplify: an antireflective trap gets you not to see the man behind the curtain by persuading you not to look at all. An inquiry trap lets you see the man behind the curtain, but tells you he's actually something else. And an insensitivity trap tells you not to care about or pay serious attention to the man behind the curtain because he's far less important than the stock market.¹³

2 The nature of playfulness

Intellectual playfulness, I suggest, is an epistemic virtue. Part of what makes it a virtue is its ability to help us escape from epistemic traps. But what, exactly, is intellectual playfulness? Let's start by taking a step back and thinking about playfulness in general. The term seems to denote a loose cluster of related qualities, which do not seem to admit of any clear and simple definition.¹⁴ But there are certain features that recur through the many discussions of play and playfulness which will serve as a useful starting point.

Let's say that *play* is a certain type of activity, and *playfulness* is the disposition to engage in play activities.¹⁵ To understand playfulness, then, we'll need to understand play. In the many discussions of play, we see two recurring qualities. First, play is done for its own sake. We play because playing is fun, pleasurable, or satisfying, and not because we want some kind of product. Second, play involves some sort of shifting of perspectives, or stepping outside of one's normal rules and roles—and stepping into other ones. Let's look at these qualities separately, before we fit them together.

First, play is *autotelic*. It is an activity engaged in for its own sake, rather than the sake of its products. We play because we want to be playing, and not because playing grants us some valuable product. Bernard Suits puts it quite nicely. In Suits' account, play is the diversion of normally instrumental resources into autotelic activity (Suits 1977).¹⁶ When we play with our food, we are taking a substance normally used for nutrition, and using it in some amusing process of stirring and sculpting. When we play wrestle, we take our physical capacities—and our fighting abilities—and use them to make a ruckus in the dirt for the raw joy of it. What matters here is the motivation for play, and not what benefits play may grant us. I may derive further benefits from play, but when I play, I am motivated by the play itself. Playful dancing may have the side-benefit of improving my fitness—but if I dance primarily for the sake of fitness, then it wouldn't be play.¹⁷

Crucially, Suits notes that “play” and “playing a game” are conceptually distinct. Games, for Suits, are particular structures of artificial

goals and voluntary obstacles. There are instances of play which are not playing a game—like playing with your food or playing with your beard. These activities involve no rules or goals. And there are instances of playing a game which are not play—like a miserable professional boxer, just doing it for the money. And there are many cases in which we are playing a game in both senses—like when we play a boardgame, exercising our intellectual capacities for the sheer fun of it, inside a structure of rules and goals.¹⁸ *Play*—which is not the same as game-play—is autotelic, in the sense that it is done for the value of being engaged in the activity of play itself, rather than for some outcome of that activity. We are playing a game as play when we are doing it for autotelic reasons. But we are playing a game as work when we are just doing it to extract some benefit, like status or money.

Second, play seems opposed, in some way, to order and strict rule-boundedness. Miguel Sicart puts it this way: true play is essentially free and appropriative. It disrupts the normal states of affairs (Sicart 2014, 3). Friedrich Schiller's account starts from a similar nubbin: play, says Schiller, is a state of openness towards the rules that normally govern you, and a willingness to transcend them.¹⁹ But play's relationship to order, rules, and norms is not merely oppositional. Play is not the same as chaos, destruction, or the refusal to follow any sort of norm whatsoever. Play often seems to involve, not just stepping away from the normal rules that guide one's life, but slipping into new ones. In the classic discussion of play, *Homo Ludens*, Johan Huizinga suggests that what it is to play is to enter a "magic circle" where we take on different roles and accept different rules. When we play a game, friends slip into the roles of enemies; mundane objects take on a special significance.²⁰

Or, as Maria Lugones puts it:

The playfulness that gives meaning to our activity includes uncertainty, but in this case the uncertainty is an *openness to surprise*. This is a particular metaphysical attitude that does not expect the world to be neatly packaged, ruly. Rules may fail to explain what we are doing. We are not self-important, we are not fixed in particular constructions of ourselves, which is part of saying that we are *open to self-construction*. We may not have rules, and when we do have rules, *there are no rules that are to us sacred*. We are not worried about competence. We are not wedded to a particular way of doing things. While playful we have not abandoned ourselves to, nor are we stuck in, any particular "world." *We are there creatively*.

(Lugones 1987, 16)

To pull a simple thread in common from all these accounts: playfulness involves a certain fluidity with respect to norms and beliefs.

It is useful here to compare playfulness with irony. To be ironic, in its most extreme form, is to refuse to value anything, or to be committed to anything—to, as Jorge Portilla puts it, enter into a complete suspension of seriousness.²¹ This refusal makes it impossible to become invested in any sort of community—since communal action requires that we commit to doing things together and valuing things together. As Søren Kierkegaard says, the pure ironist wants to be entirely free from obligations, restrictions, and commitments; this dedication to pure negative freedom makes them unable to participate substantially in much of human life (Frazier 2004, 419–421).

But play is different. Play involves lightness with rules, in both directions—the ability to lightly step away from, but also the ability to lightly adopt. Think about the difference between playfulness, seriousness, and irony, when it comes to playing games. To be serious about a game is to play it under the idea that its goals are really and genuinely important—as, say, an Olympic athlete does. The opposite of such seriousness is the wholly ironic game player. They refuse to adopt any of its norms in any committed way. But that sort of irony is often antithetical to the shared commitment of game-play. Such an attitude, Huizinga says, makes one a *spoilsport*, who mocks the game and wrecks the shared illusion of gameplay (Huizinga 1980, 11). To be playful about games is neither to be utterly serious or utterly ironic, but to move easily into and out of commitments to rule-sets.

Consider, for example, the shared mood of tabletop roleplaying games. The players have to commit, temporarily, to the rules of the game and a kind of (absurd) sincerity of purpose. The players have to really go all-in in pretending to be in character—of really being, say, fantasy elves and dwarves on a quest to save a village. As is often remarked by dedicated role-players, this shared mood is often wrecked by the pure ironist—who mocks the activity, who follows the rules mechanically but without real commitment, who breaks the illusion by calling attention to the arbitrariness of its rules (Nguyen 2019). Francisco Gallegos makes a parallel point in his discussion of Portillian irony. So much of human life, says Gallegos, depends on a shared mood. But such moods are delicate and require considerable communal support. They depend, one might say, on creating resonance through active participation. An ironist, by openly refusing that shared commitment, destroys the communal development of shared moods (Gallegos 2013, 13–14).

So playfulness involves not only the ability to slip away from one framework of norms and beliefs, but also the ability to slip into a new framework—at least for a while. To be playful with a game is to bring oneself to care, for a time, about the specified goals of the game, and to adopt, for a time, a temporary but absolute obedience to a set of rules. It involves entering into, in some phenomenally substantial way, the imagined world of the game. And it involves letting those goals and rules slip

away when the game is done. To be playful with a game is to wear the game's cares and norms lightly (Nguyen 2019; 2020a, 27–73, 216–224). The ironist may mock, but they don't have quite the same spirit of light-heartedness. They wear their refusal to participate too heavily, to play.

If we were interested in constructing an account of playfulness in general, things would turn much more complicated around this point. But I think we have enough bits and pieces gathered to make a stab at saying something about the narrower quality of intellectual playfulness. Intellectual playfulness seems to include the ability and interest in trying out new ideas, perspectives, and belief systems. Let's call this the disposition for *perspective shifting*. The playful person can step out of a framework of beliefs, values and cognitive framing mechanisms, and step into another. Those new perspectives may be only temporary visiting points, or they may grow into something that the person inhabits more deeply. The playful person is neither dogmatist nor ironist, but, as Lugones puts it, an easy traveler between, and an explorer of, different normative worlds.

Let's put our two parts together, now. I propose that *intellectual playfulness* is the disposition to investigate ideas, beliefs and values in a manner that is:

- 1 *Autotelic*—done for the sake of being involved in the investigation itself and
- 2 Involves intellectual *perspective shifting*—trying on and (at least temporarily) inhabiting alternate belief systems, which includes trying out alternate beliefs, values and norms for belief-acquisition.

In shorthand: intellectual playfulness is the disposition to try out new perspectives for fun. For brevity's sake, I'll refer to intellectual playfulness as “playfulness” for the remainder of this chapter—but where it should be understood that I am not attempting to speak about the whole vast edifice of playfulness in all its ineffable glory, but only about this specific cognitive varietal.

3 The value of perspective shifting

Why would this form of playfulness be an intellectual virtue? A disposition to engage in perspective-shifting seems obviously valuable for epistemic pursuits. But why might it be especially virtuous to do it for fun? Before we answer the complex question about fun, let's get clear on the cognitive value of perspective-shifting.

Compare the disposition to shift perspectives with a nearby neighbor: the attitude of open-mindedness. Open-mindedness is a disposition to be open, to a certain extent, to challenges to one's own beliefs, taking them seriously rather than dismissing them. Wayne Riggs' account offers us a useful way to flesh out of this notion. There's a difficulty, says Riggs,

for any philosophical accounting of open-mindedness: open-mindedness seems incompatible with full-throated belief. Why should we seriously consider challenges to a particular belief, if we were already confident in that belief? Riggs solution is to take open-mindedness to be, not an attitude towards particular beliefs, but rather an attitude one holds towards oneself as a believer, in general. Open-mindedness involves a general awareness of one's fallibility as a believer, and the general acknowledgment that for any belief, one might be wrong (Riggs 2010, 180).

Riggs points out that being open-minded doesn't require us to take seriously every single challenge to our beliefs. (That would open the door to an overwhelming cognitive load, for one thing.) Rather, open-mindedness involves using our positive knowledge of our likely fallibilities to decide which challenges to take seriously. Suppose there were a bunch of musical artists that I think are just crap. (You don't really have to suppose it—it's true.) I might not take seriously each and every challenge to my musical judgments. The fact that Smashmouth has legions of loving fans doesn't, by itself, give me any reason to relisten to those horrible Smashmouth singles. But suppose that my friend points out that the overall pattern of my musical judgments reveals a systematic bias: I seem to reliably prefer white artists over black artists. This claim hooks up with my background knowledge about the state of the world—about my having grown up in a systematically prejudiced society. My positive understanding of my potential for fallibility gives me reason to take a particular set of challenges seriously—like, say, my dismissal of rap.²²

Open-mindedness, then, turns out to be quite different from perspective-shifting. Open-mindedness makes a weaker demand. An open-minded person ought to take some challenges seriously, when their background belief system gives them good reason to. But their standing belief system is a very active participant in the process. First, their belief system shapes which challenges they take seriously. Second, when they do take a challenge seriously, that challenge will be investigated using their standing belief system. Open-mindedness is a willingness to entertain challenges *when those challenges are properly supported by other parts of one's current belief-system, where the ensuing investigation will be conducted using one's current belief system*. Open-mindedness is a good guard against the possibility that my belief-system has not been made adequately coherent. I might have formed my judgments of musical artists based on my immediate response of pleasure, and never connected that up with my background beliefs about bias—until somebody else challenged me to.

But open-mindedness, understood this way, is particularly weak against epistemic traps, especially inquiry traps. In an inquiry trap, beliefs come in a self-supporting network, which contains resources to repel challenges. When you are open-minded, you are willing to consider challenges. But the process of inquiry, for the open-minded person,

draws upon their background beliefs—and, in an inquiry trap, those background beliefs function to re-assert the original belief system, and offer explanations and considerations to block challengers. Mere open-mindedness leads us to inquiries conducted while using our standing belief-system. And in an inquiry trap, that belief system has been rigged to reaffirm itself. If the function of open-mindedness is to iron out incoherencies in one's belief system, then it won't help against a trap belief system that has already been engineered for appealing internal consistency.

Perspective-shifting, on the other hand, involves actively trying on a new perspective. It involves going through—or at least, entertaining—lines of inquiry from alternative systems of belief. The perspective shifter will not only re-consider a single belief or narrow set of beliefs, but also be willing to consider it from the perspective of a temporarily-adopted alternate belief-system. The value of perspective-shifting is in its temporary suspension of one's standard belief system. Perspective-shifting is an insurance policy against inquiry traps because it can neutralize, for a time, those engineered, pre-prepared defenses. Perspective-shifting gives alternate belief system some air, so that the shifter can explore an alternative system of explanation as a functioning and networked whole—rather than rejecting the parts piecemeal, from the perspective of their standing belief-system.²³

Let me offer an analogy, in the key of Otto Neurath. Imagine that a belief system is a boat. Open-mindedness involves the willingness to pull out any particular plank and inspect it, to see if it's really the best plank for the job. But that assessment occurs while standing on all the other planks of that boat. Each particular evaluation of each particular plank will still occur against the background of the rest of the planks. So even if you assess each and every plank individually, the boat will retain its shape. Perspective shifting involves jumping ship and trying out a whole new boat.

4 The cognitive value of fun

But perspective-shifting is not, by itself, playfulness. Playfulness involves engaging in perspective-shifting activity for autotelic reasons: for the sheer fun and joy of it, for the beauty of the ideas. Playfulness can even involve delight in the perspective-shifting itself—in the joys of trying to occupy a particularly strange and alien position. We can relish a new mental position for its mind-bending weirdness—for the delightful feeling of having to stretch our minds into some odd shape.

So here is the key question: why might perspective-shifting be epistemically better when it is done autotelically, rather than when it is used as an instrument for the pursuit of epistemic ends?

The question might seem quite strange at first. How could the fun-loving idea-player ever get closer to the truth than somebody who was

directly pursuing the truth? But the idea is not entirely outlandish. What we are approaching here is the possibility that truth might be somewhat related to what are called “self-effacing ends”. A self-effacing end is an end that cannot be acquired through direct pursuit. A classic example is *the pleasures of love*. There are certain pleasures associated with loving another person—with being unselfishly devoted to promoting another’s interests. But an entirely selfish person couldn’t get the pleasures of love. If a pure egoist were trying to be in love, for the sake of their own selfish enjoyment of the associated pleasures, then they wouldn’t actually be in love. The pleasures of love are self-effacing (Parfit 1984, 23–24; Pettigrove 2011, 192–193; Nguyen 2020a, 53–58). Similarly, the playful person might have an advantage in getting certain epistemic goods, if it turned out that those epistemic goods were self-effacing, at least in part.

Why might that be? I think there are two distinct, but interrelated possibilities. First, an interest in getting it right constrains the search space, focusing searches on areas that promise good epistemic yields. Suppose that you are perspective-shifting, not for autotelic reasons, but in the pursuit of truth. You are searching the possibility space for ideas you might have missed. Your perspective shifts will likely be guided by your sense of which shifts will be epistemically fruitful. Since you are interested in the truth, you’ll try on those alternate systems of belief which might turn out to be true. Your shifts will be constrained by your sense of plausibility. And that assessment will proceed from your standing system of beliefs. Even if you are trying out alternative systems of belief, the choice of those systems will still be influenced by your standing system of beliefs.

But a well-designed epistemic trap should be able to manipulate these plausibility assessments. A well-designed inquiry trap can undermine the plausibility of key alternate perspectives by, for example, associating them with the most wildly untrustworthy and unsavory people. I take it that you or I would probably never even attempt to occupy the moral perspective of, say, a Nazi, as part of a search procedure for real moral truth. An echo chamber could strategically manipulate that effect, by associating alternative moral and political visions with that kind of sheer outright evil, as part of their strategy of credential manipulation. Jamieson and Cappella note that one of Rush Limbaugh’s basic strategies for building his echo chamber is creating an insider language full of emotionally-charged labels for opponents and their positions. This language serves both to create an “insular language community”, and to reinforce associating outsider belief systems with pure evil. For example, Limbaugh coined the term “feminazi”, which strongly associates the position of feminism with fascism, putting it beyond the moral pale (Jamieson and Cappella 2010, 177–190). Our analysis here makes Limbaugh’s maneuver clear: he is trying to make feminism seem so wildly implausible, as to be unworthy of any exploratory efforts.

But somebody who was perspective-shifting for autotelic reasons—for the fun of it, for the beauty of the ideas, for the joy of the sheer perspective-shifting itself—would be freed from those plausibility constraints. They don't need to engage their standing background beliefs to figure out which alternative perspectives to occupy, since their reason for occupying alternative perspectives has nothing to do with those perspectives' likely truth. Playfulness is unconcerned with truth, and so unconcerned with plausibility—and so freed from such dismissals emanating from background beliefs. Playfulness, as a motive, brings people to explore belief systems that their current background beliefs treat as beyond the pale.

This, of course, has its dangers. But it also has a clear functionality: it provides an insurance policy against epistemic traps. This is not as implausible as it might seem. What this looks like, in actual life, is people trying out and exploring systems of belief because they are funny, beautiful, elegant, or charmingly bizarre. In my own life as a teacher, I've noticed that these sorts of motivations often get students to let down their guard for a moment. When I present certain philosophical theories as candidates for the truth, when those theories are sufficiently distant from my students' own belief system, my students are likely to reject them immediately, without significant consideration. But when I present philosophical theories as worth thinking about because they are gorgeously elegant or deliciously fun, then students will actually try them out for a while—and often find that these belief systems can carry more water than it had first seemed.

Another way to put the same point: rational beings need to go on some random walks. It is easy, says Adrian Currie, to get trapped in local maxima during the inquiry process. Attempts to optimize for truth will help climb a local maximum, but are likely to miss higher peaks that are radically different. So the right thing to do is to sometimes go on *random walks*—to explore ideas unconstrained by the need to optimize for truth every step along the way. And, he says, we have a name for the tendency to go on a reasonable number of random walks: we call it “creativity” (Currie 2019). As Sara Aronowitz says, the optimally rational being—or community of beings—mostly pursues the best-looking most plausible paths for exploration, but occasionally goes on random walks (Aronowitz 2021).

Of course, one might simply protest: if going on occasional random walks—and occasionally occupying implausible perspectives—is part of the best path for rationality, then shouldn't the rational person simply make themselves go on random walks? Surely a rational person should think that this would be the right strategy. But what would this actually look like, as a plausible activity that could be adequately motivated in a human? It seems difficult to imagine that a person interested only in the truth would be adequately motivated to explore, carefully and

thoroughly, a completely implausible position. If we wanted to construct a rational being with cognitive limitations, who occasionally went on random walks with some degree of care, then we should build a being that *enjoyed* sometimes going on random walks, with no thought that they would take them somewhere good. As David Schmidtz says, an agent that loves eating and sex for their own sake will do better at survival and procreation than an agent who values survival and procreation, and pursues eating and sex only as instruments to those final ends (Schmidtz 2001, 251–255). Intellectual playfulness can directly motivate epistemic agents to explore the space of possibilities, sometimes leaving behind considerations of plausibility. (Autotelicity isn't the only possible motivation, however. We can easily imagine others. For example, we might set up an institution that strongly incentivized the publication of ideas merely because they were novel, and not because they were likely to be true. This would also incentivize people to explore the possibility-space, away from plausibility constraints.)

One might worry that playfulness is just as likely to get a person ensnared in a new epistemic trap as it is to get them out of one. After all, might one not explore an epistemic trap and so become seduced by it, in the exploration? This is certainly a possibility. But one thing we might say is that playfulness serves as a useful insurance policy when it occurs in epistemic agents that are otherwise mostly rational. That is, a rational epistemic agent should be able to, if adequately presented with two systems of belief, determine which is better. Epistemic traps work to keep rational people in epistemically inferior systems of belief by preventing them from getting an adequate view of the alternatives. So, for such a rational epistemic agent, playful exploration of the space will get them out of epistemic traps. But for an irrational epistemic agent, easily seduced by, say, clear-seeming explanations, playfulness may get them into trouble. This is just to say that playfulness won't get us to intellectual virtue by itself. It is useful as a motive to explore widely, but that exploration will only bear fruit when appropriately conjoined with other intellectual virtues.

Importantly, playfulness suffers from its own particular form of constraints. The hedonistically-motivated form of playfulness I've described is not entirely free-ranging. It will tend to seek out and linger on those belief systems that give us some kind of pleasure—the beautiful ones, the fun ones, the entertainingly wild ones. That is why, I suspect, a really robust epistemic character will involve multiple dispositions to shift perspectives for different reasons. Consider, for example, empathy. Empathy, some have suggested, is the disposition to take on the emotional perspective of another person.²⁴ But notice that empathy, too, has its weaknesses and vulnerabilities. We might only be empathetic to people that we spend significant time with, or those we think are worthwhile people. And epistemic traps can manipulate those qualities too. A

well-constructed echo chamber, for example, can bring you to limit the amount of time you spend around outsiders, and also treat those outsiders as monsters beyond the moral pale.

It will be most useful, then, to maintain a variety of different perspective-shifting dispositions, each of which perspective-shifts for different reasons, and each of which has its own vulnerabilities. Truth-oriented perspective shifting is limited by one's sense of plausibility; playfulness is limited by one's pleasure; empathy is limited by one's social sphere. We need a diverse portfolio of perspective-shifting dispositions, each of which will do some work to shore up the limitations of the others.

To sum up: Playfulness brings us to explore other perspectives. It provides the motivational force to leave well-ordered belief systems and explore new ones. And that is particularly useful against epistemic traps. In many cases, the belief system of an epistemic trap would be, to the eyes of a genuinely rational agent, obviously worse than other belief systems. The trap works on such agents by occluding those alternative belief systems, so an adequate comparison can never be made. The trap can't usually completely block out those alternative belief systems from view. They can work, instead, by keeping entrapped agents from spending time exploring those alternative belief systems—which they can often do by presenting such exploration as worthless or silly. Playfulness is a disposition that provides the motivation to explore alternative belief systems, coupled with the technique of suppressing one's background beliefs. It seems precisely tuned to block the workings of this sort of epistemic trap.

5 Pleasurable attention

Autotelicity has a second important function, besides freeing us from plausibility constraints. My discussion here will depend on an empirical claim about our psychology, though one with significant empirical support.²⁵ Suppose, for the moment, that pleasure attracts our attention. We attend to that which we enjoy and care about. When we love the process of doing something, we pay more attention to the details of that process, than if the process were a mere instrument.

This relationship was made clearest to me when I was learning to rock climb. As a novice, I was driven by the desire to get to the top, flinging myself at the wall in earnest efforts. A friend—and a far better climber—told me: “Just savor your movement, OK? Just love the motion”. At first, I thought this was strictly a comment about the value of the activity—and, indeed, it did make rock climbing a far richer and more lovely experience. But, interestingly, the more I let myself focus on the pleasures of movements, the better a climber I became. This is, I take it, because the attitude of taking pleasure in my movement drives me to attend more lovingly to every aspect of my movement, to take in

the details. And for hard rock climbing, the climber needs careful control of the subtle details of their movement. The activity of savoring my movement for its own sake, then, also supports the development of my sensitivity toward my own body and its movements. For similar reasons, those cooks who love the process of cooking tend to turn out much better food, in the end, than those cooks who are interested primarily in the end-product. Pleasure is not the only way to drive attention somewhere; we can also force our attention there, through sheer effort of will. But a being constituted to take pleasure in the process of doing something will need to spend far less emotional and cognitive resources to get themselves to attend to the details of that process, than a being who finds such attentions unpleasant, but exerts them through force of will.

What's more, if we take pleasure in attending to a process for its own sake, we will likely see the details of the process more clearly. Why might this be? There's a useful lead in aesthetic theory, in a discussion about the special nature of aesthetic attention and perception. Consider the aesthetic attitude. According to one traditional line of thinking, the aesthetic attitude is quite a distinct one from the everyday practical attitude. In ordinary life, we have clear practical goals, and we look to the objects in our lives to meet those practical goals. Our attention is filtered: we pay attention to those features of the object relevant to our practical interests, but not the irrelevant features. If we need a hammer just to hammer in some nails, we would pay attention to its weight, balance, heft, and hardness—but not to the color of the wood, the smell of the iron, the pattern of patina on the rust. Our attention, when it is practical, is narrowed and specific. But when we attend aesthetically, we do so for the value of the experience of attending itself. And so our attention roves over all parts of the object in an unfiltered way.²⁶

Though the argument is couched in terms of the “aesthetic”, the argument relies on one particular feature of the aesthetic attitude: that it is marked by the attitude of disinterestedness. In the aesthetic attitude, we attend to an object for its own sake, rather than for the sake of using that object as an instrument to some other end. The argument actually works, then, for any autotelic form of attention. Playful attention is just as disinterested as aesthetic attention, and so just as unfiltered.

If we have an object that we consider under one single use-category, and we only look to it with an eye towards that use, then we can easily fail to notice other aspects, that might make it useful in other ways. So long as I look at this whisk for cooking, I will only pay attention to its practically relevant features—the grip on the handle, whether it has the right shape for beating eggs, etc. Only when I take an unfiltered, aesthetic attention will I also notice the pleasingly eccentric noise it makes when struck, and the delightful way it shivers in unpredictable self-clattering loops. And those kinds of observations might let me see new uses for it—like, for instance, that the whisk also turns out to be an

absolutely magnetic toy for babies to pound things with. The creative use of objects, then, involves a touch of self-effacement. The person who is esthetically interested in the object may have an advantage in seeing the object in all its totality—a process that may reveal new and unexpected uses for the object. This means, paradoxically, that the esthetic attitude is quite useful—and useful precisely because it is unconcerned with the usefulness of its object.

The same, I think, is true for ideas and belief systems. When we assess a belief system for its usefulness to us, our vision narrows. Let me start with an extreme—but familiar—case. Suppose we have made up our minds about some issue. Our interest in arguments towards those issues will typically be practical—we may be interested in using them to convince other people, or to fend off attacks and criticisms. We will attend to those features that are useful for that end. We are unlikely, then, to explore in detail the way an argument works that carries us to some other target. (And if we do, we will likely be paying closest attention to where we might find flaws.) But if we try it on in a spirit of play, then that practical filter is lifted. We can explore how the argument works—the way a belief system coheres—in an unfiltered way. And the more pleasure we take in it, the more we will attend to the details—discovering new possibilities that we might not have seen before.

We can find a subtler version of that effect in less extreme cases. When I attend to ideas in the mode of truth-seeking, I notice the features of those ideas which strike me as useful in the pursuit of truth. The selection of those features will, again, be driven by my sense of the plausible. But in playful exploration, we don't confront ideas by immediately assessing them for their usefulness—so we can linger on the details of a stranger belief-system.

Such open and unfiltered attention is an antidote to epistemic traps that function by directing attention away from relevant alternatives. Such attention seems particularly potent against inquiry traps and insensitivity traps. In an inquiry trap, a belief system manipulates plausibility considerations so as to prevent the believer from lingering in what are genuinely good, alternative belief systems. In an insensitivity trap, a narrowed sense of what is really valuable sharply focuses the attention and shrouds other domains beneath a veil of unimportance. A belief system needs to be given some time and energy, before its powers become apparent. In each case, some valuable alternative is choked of air.

Playfulness motivates people to spend some time in alternative belief systems, unconstrained by the limitations of their background belief system. Playfulness gives the entrapped person some reason to explore unimportant-seeming domains, to reason from within those alternate perspectives. Playfulness motivates people to try out ideas, not because they are plausible or important, but because they are fun and beautiful. And those qualities are, if not entirely random, at least importantly skew

of how our usual epistemic goals, values, and beliefs guide us—and so free of the traps that might have been built into our standing set of goals and beliefs. The claim here is not that we should always be animated only by a sense of fun in our intellectual life. It is that playfulness is an excellent attitude to occasionally take up—that will drive us out of our usual intellectual paths, and encourage us to occasionally leap into a faraway perspective.

Of course, if we wanted to engineer an effective epistemic trap, then we will want to discourage playfulness. We will want to cultivate a kind of bloody serious-mindedness, a disdain for intellectual play for play's sake. And this is what I think we often find, in real-world epistemic traps: the spirit of playfulness is discouraged—labeled as evil or wasteful. Playfulness is particularly easy to exclude in insensitivity traps. We simply need to articulate the values of an insensitivity trap in a way that leaves playfulness by the wayside. (For example: valuing strictly money, or valuing strictly rigor.) This gives those of us, who are opposed to epistemic traps, a reason to try to cultivate and spread the virtue of playfulness as an antidote.

Acknowledgments

I'd like to thank Elizabeth Camp, Anthony Cross, Adrian Currie, Nick Hughes, and Melissa Hughes for their wise counsel for this chapter.

Notes

- 1 I am being intentionally vague here between about what counts as a good beliefs. I am trying here to make no particular commitments about epistemic internalism vs externalism, reliabilism, pragmatism, or to take sides on any of the major epistemic debates of the contemporary scene. In particular, when I say that good-faith epistemic efforts are those that proceed from epistemic grounds. In particular, I mean for my account here to include, as good-faith beliefs, those beliefs not supported by evidence, but whose adoption supports epistemic goals. My hope here is that I can give an account of epistemic traps compatible with any of the standard positions of contemporary epistemology.
- 2 Elsewhere, I've discussed the possibility that some belief systems offer us a hedonistic instrumentalization, by giving us pleasure in return for adopting certain belief systems (Nguyen and Williams 2020; Nguyen 2021a, 2021b).
- 3 Joshua DiPaolo offers a useful study of the epistemic manipulations of fanaticism—which often involve undermining followers' self-trust as well as their trust in outsiders, and placing that trust entirely in the hands of a small leadership (DiPaolo 2020).
- 4 The idea of, and name for, “apathy traps” suggested by Geoff Pynn.
- 5 Geoff Pynn suggested the terms “apathy trap” and “inquiry trap”, and greatly assisted in the development of this taxonomy.
- 6 The ensuing paragraphs offer a brief summary of my analysis of echo chambers in (Nguyen 2018).

- 7 For a more recent discussion, see *Network Propaganda* (Benkler et al. 2018). I offer a discussion of their view in (Nguyen 2021c).
- 8 See also (Nguyen 2020b) for a discussion of a more minimal kind of non-engineered epistemic trap—one in which erroneous beliefs lead to the selection of unreliable experts, which reinforces those erroneous beliefs.
- 9 For more on the satisfactions of simple explanations offered by echo chambers, see Nguyen (2021b).
- 10 This example adapted from Jamieson and Cappella’s analysis of Rush Limbaugh’s rhetorical strategies.
- 11 Note, however, that merely because something is a conspiracy theory doesn’t mean that it is false, or that its believer is in an inquiry trap. There are, after all, real conspiracies in the world, and rational people should believe in some conspiracy theories (Coady 2012, 110–137; Dentith 2019). But conspiracy theories can function as part of a well-tuned strategically formulated inquiry trap.
- 12 This account of the insensitivity trap is only a brief sketch, and the description of this businessperson something of a cartoon; I plan to develop this account in future work.
- 13 I owe this analogy to Melissa Hughs.
- 14 For an argument to this effect, see Randolph Feezell’s argument the concept of “play” is essentially pluralistic, and none of the main categories can be reduced to another (Feezell 2010). For an anthropologist’s discussion to a similar effect, see Brian Sutton-Smith’s famous account of the ambiguity of play (Sutton-Smith 2001).
- 15 There is some debate about which of these concepts is primary and which secondary. For example, Bernard Suits thinks that “play” is primary, Maria Lugones thinks that “playfulness” is primary (Suits 1977; Lugones 1987). My analysis attempts to remain agnostic on that debate.
- 16 For a useful exploration and refinement of the details of Suits’ position, see Emily Ryall (2013).
- 17 This kind of strictly motivational account is an improvement of an earlier, more demanding sort, like Johan Huizinga’s, which specified that play both proceed from no interest in benefit, and actually grant us no benefit (Huizinga 1980, 1–20). But so many paradigmatic instances of play obviously offer benefits in physical fitness and mental health, among other things.
- 18 (Suits 1977). For Suits’ account of games as constructs of artificial goals and constraints, see (Suits and Hurka 2014).
- 19 This is a vast oversimplification of a very complex theory. For a detailed discussion of Schiller’s theory of the play drive, how it unites the rational and the sensual, and how it provides an account of aesthetic value, see Samantha Matherne and Nick Riggle’s reading of Schiller’s *Letters on the Aesthetic Education of Man* (Matherne and Riggle 2020).
- 20 This suggestion leads to a rather vast literature on what’s called “the magic circle” — the alternate space of play. There have been some significant criticisms of the concept (Taylor 2007; Malaby 2016). But I think modern reconstructions of the view are much more plausible (Stenros 2012; Waern 2012). I have offered my own reconstruction and defense of the magic circle concept (Nguyen 2020a, 177–180).
- 21 My understanding of Portilla is shaped by discussions by Carlos Alberto Sánchez (2012) and Francisco Gallegos (2013).
- 22 This actually happened to me as a college freshman. Taking my friend’s challenge seriously led to the greatest internal aesthetic revolution of my life—and the most valuable one. I offer a detailed discussion of trust and prejudice in esthetic appreciation in (Nguyen 2021d).

- 23 Some may wish to call perspective-shifting a kind of open-mindedness. The precise terms here don't seem particularly important to me. What seems important, rather, is the difference between the two attitudes, and the difference between the willingness to consider challenges and perspective shifting. We could just as easily call the attitude described by Riggs "weak open-mindedness", and call perspective-shifting "strong open-mindedness".
- 24 Peter Goldie offers a useful summary of some accounts of empathy as perspective shifting (Goldie 2011). Goldie also offers a criticism: he thinks full-blooded empathetic perspectival-shifting is impossible, because you will never really be able to see something fully from another's emotional perspective—some version of you will always come along for the ride. He permits weaker forms of perspective-shifting—where you imagine what you would think while adopting some aspects of another's perspective. That weaker form is all we need for these trap-escaping qualities. You don't need to take on every aspect of another's personality to explore an alternate belief-system.
- 25 For an overview of this empirical support, and a plausible application to understanding how pleasure motivates and facilitates aesthetic appreciation, see Matthen (2017).
- 26 The aesthetic attitude thesis is usually attributed to Jerome Stolnitz (1960). Thought it became unpopular through some supposedly decisive counterarguments from George Dickie (1964), the argument has seen plausible contemporary defenders (Kemp 1999). Most notably, Bence Nanay has offered an empirically-informed account of esthetic perception, based in contemporary research into the cognitive psychology of perception, which supports a revised version of Stolnitz's esthetic attitude thesis (Nanay 2016, 1–34).

References

- Aronowitz, Sara. 2021. Exploring by Believing. *Philosophical Review*, 130(3): 339–383.
- Begby, E. (2020). Evidential Pre-emption. *Philosophy and Phenomenological Research*, 102, 515–530. <https://doi.org/10.1111/phpr.12654>
- Benkler, Yochai, Robert Faris, & Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford: Oxford University Press.
- Coady, D. (2012). *What to Believe Now: Applying Epistemology to Contemporary Issues*. West Sussex: Wiley-Blackwell.
- Currie, A. (2019). Existential Risk, Creativity & Well-Adapted Science. *Studies in History and Philosophy of Science Part A*, 76, 39–48. <https://doi.org/10.1016/j.shpsa.2018.09.008>
- Dentith, M. R. X. (2019). Conspiracy Theories on the Basis of the Evidence. *Synthese*, 196(6), 2243–2261. <https://doi.org/10.1007/s11229-017-1532-7>
- Dickie, G. (1964). The Myth of the Aesthetic Attitude. *American Philosophical Quarterly*, 1(1), 56–65.
- DiPaolo, J. (2020). *The Fragile Epistemology of Fanaticism* (pp. 217–235). London: Routledge. <https://doi.org/10.4324/9780429325328-11>
- Feezell, R. (2010). A Pluralist Conception of Play. *Journal of the Philosophy of Sport*, 37(2), 147–165. <https://doi.org/10.1080/00948705.2010.9714773>
- Frazier, B. (2004). Kierkegaard on the Problems of Pure Irony. *Journal of Religious Ethics*, 32(3), 417–447. <https://doi.org/10.1111/j.1467-9795.2004.00173.x>

- Gallegos, F. (2013). Seriousness, Irony, and Cultural Politics: A Defense of Jorge Portilla. *American Philosophical Association Newsletter on Hispanic/Latino Issues in Philosophy*, 13(1), 11–18.
- Goldie, P. (2011). Anti-Empathy. In A. Coplan & P. Goldie (Eds.), *Empathy: Philosophical and Psychological Perspectives* (p. 302). Oxford: Oxford University Press.
- Huizinga, J. (1980). *Homo Ludens: A Study of the Play-Element in Culture*. London: Routledge & K. Paul.
- Jamieson, K. H., & Cappella, J. (2010). *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford: Oxford University Press.
- Kemp, G. (1999). The Aesthetic Attitude. *British Journal of Aesthetics*, 39(4), 392–399. <https://doi.org/10.1093/bjaesthetics/39.4.392>
- Lugones, M. (1987). Playfulness, “World”-Travelling, and Loving Perception. *Hypatia*, 2(2), 3–19. JSTOR.
- Malaby, T. M. (2016). Beyond Play: A New Approach to Games. *Games and Culture*, 2(2), 95–113. <https://doi.org/10.1177/1555412007299434>
- Matherne, S., & Riggle, N. (2020). Schiller on Freedom and Aesthetic Value: Part I. *British Journal of Aesthetics*, 60(4), 375–402. <https://doi.org/10.1093/aesthj/ayaa006>
- Matthen, M. (2017). The Pleasure of Art. *Australasian Philosophical Review*, 1(1), 6–28. <https://doi.org/10.1080/24740500.2017.1287034>
- Nanay, B. (2016). *Aesthetics as Philosophy of Perception*. Oxford: Oxford University Press.
- Nguyen, C. T. (2018). Echo Chambers and Epistemic Bubbles. *Episteme*, 17, 1–21. <https://doi.org/10.1017/epi.2018.32>
- Nguyen, C. T. (2019). The Forms and Fluidity of Game-Play. In T. Hurka (Ed.), *Games, Sports, and Play: Philosophical Essays*. Oxford: Oxford University Press. <https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198798354.001.0001/oso-9780198798354-chapter-4>
- Nguyen, C. T. (2020a). *Games: Agency as Art*. Oxford: Oxford University Press.
- Nguyen, C. T. (2020b). Cognitive Islands and Runaway Echo Chambers. *Synthese*, 197(7), 2803–2821.
- Nguyen, C. T. (2021a). How Twitter Gamifies Communication. In Jennifer Lackey (Ed.), *Applied Epistemology* (410–436). Oxford: Oxford University Press.
- Nguyen, C. T. (2021b). The Seductions of Clarity. *Royal Institute of Philosophy Supplements*, 89, 227–255.
- Nguyen, C. T. (2021c). Polarization or Propaganda? *Boston Review*. April 22, 2021. <http://bostonreview.net/politics-philosophy-religion/c-thi-nguyen-polarization-or-propaganda>
- Nguyen C. T. (2021d). Trust and Sincerity in Art. *Ergo* 8(2): 21–53.
- Nguyen, C. T., & Williams, B. (2020). Moral Outrage Porn. *Journal of Ethics and Social Philosophy* 18(2): 147–172.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Pettigrove, G. (2011). Is Virtue Ethics Self-Effacing? *The Journal of Ethics*, 15(3), 191–207. <https://doi.org/10.1007/s10892-010-9089-4>
- Riggs, W. (2010). Open-Mindedness. *Metaphilosophy*, 41(1–2), 172–188. <https://doi.org/10.1111/j.1467-9973.2009.01625.x>

- Ryall, E. (2013). 3 Playing with Words. In E. Ryall (Ed.), *The Philosophy of Play* (p. 44). London: Routledge.
- Sánchez, C. A. (2012). *The Suspension of Seriousness: On the Phenomenology of Jorge Portilla, with a Translation of Fenomenología Del Relajo*. New York: State University of New York Press.
- Schmidtz, D. (2001). Choosing Ends. In E. Millgram (Ed.), *Varieties of Practical Reasoning* (pp. 237–257). Cambridge, MA: MIT Press.
- Sicart, M. (2014). *Play Matters*. Cambridge, MA: MIT Press. <https://mitpress.mit.edu/books/play-matters>
- Stenos, J. (2012). *In Defence of a Magic Circle: The Social and Mental Boundaries of Play*. <http://www.digra.org/wp-content/uploads/digital-library/12168.43543.pdf>
- Stolnitz, Jerome. 1960. *Aesthetics and the Philosophy of Art Criticism: A Critical Introduction*. New York: Houghton Mifflin.
- Suits, B. (1977). Words on Play. *Journal of the Philosophy of Sport*, 4(1), 117–131. <https://doi.org/10.1080/00948705.1977.10654132>
- Suits, B., & Hurka, T. (2014). *The Grasshopper: Games, Life and Utopia* (Third edition). Peterborough, Canada: Broadview Press.
- Sutton-Smith, B. (2001). *The Ambiguity of Play*. Cambridge, MA: Harvard University Press. <https://www.hup.harvard.edu/catalog.php?isbn=9780674005815>
- Taylor, T. L. (2007). Pushing the Borders: Player Participation and Game Culture. In J. Karaganis (Ed.), *Structures of Participation in Digital Culture* (pp. 112–132). New York: Social Science Research Council. <http://content.talisaspire.com/auckland/bundles/5a7a6e56540a267f7849b434>
- Waern, A. (2012). Framing Games. *DiGRA Nordic '12: Proceedings of 2012 International DiGRA Nordic Conference*, 10. <http://www.digra.org/wp-content/uploads/digital-library/12168.20295.pdf>

9b Commentary from Ian James Kidd

Comments on C. Thi Nguyen, “Playfulness Versus Epistemic Traps”

Molière, the French playwright best known for his 1666 play, *La misanthrope*, once opined that the “function of comedy is to correct the vices of men” (Molière 2001: xiv). Comedy can be effective since a common characteristic of many vicious people is that they cannot bear to be teased and cannot invest their own conduct with a sense of levity. In his excellent chapter, C. Thi Nguyen explores the role of playfulness in epistemic life by construing it as an epistemic virtue, specifically as a guard against “epistemic traps” and related forms of bad doxastic behavior. By way of endorsement, I want to quickly emphasize three aspects of his discussion, then point to some issues about the limits of playfulness as it relates to other epistemic virtues and the deeper normative commitments of the epistemic agent.

Nguyen, first, notes that the epistemically virtuous agent we meet in the pages of virtue epistemological writings is a rather serious person. In two of the earliest monographs on epistemic virtue, by Lorraine Code and James Montmarquet, the core concepts were those of epistemic responsibility and epistemic conscientiousness. “Responsible” and “conscientious” are good qualities to have, no doubt, but they hardly suggest someone who’s the life and soul of the party. Construed properly, playfulness can inject some vitality and spontaneity into our epistemic lives. Second, Nguyen emphasizes the cognitive functions of a specifically epistemic sort of playfulness, like the dispositions to “try out new ideas, perspectives and systems of thought for the sheer joy of it” and, less fun, the resistance it can generate to certain forms of dogmatism (one thinks, here, of one of José Medina’s favorite metaphors, that of “epistemic friction”). An epistemically playful person gets cognitive work done, has a good time of it, and also avoids the sorts of problematic attitudes and behaviors, like dogmatic quarrelling, that ruin the game for everyone. A third aspect of the account is the ways that playfulness serves to transform what some now call our cognitive phenomenology—roughly, our experience of thinking, reasoning, understanding and the like. Nguyen,

for instance, talks of how the playful enquirer is more open to surprise, thrills in the unexpected, enjoys working within systems of rules without being shackled by them, and so on.

What we get is therefore a richer and, in a sense, more humane vision of what a really epistemically virtuous person can be—not a conscientious plodder who joins the dots, but an individual with a sense of *élan*, warm to excitement and able to enjoy the work of the mind without lapsing into an obstructive seriousness that dampens the mood for others as well as increasing one's susceptibility to patterns of mental inflexibility. Nguyen is rightly esteemed for his insights into the philosophical seriousness of human ludic practices—of games and the whole world of play—and this chapter is a case in point.

I want to flag, though, two sets of issues that concern the limits of the sorts of virtuous epistemic playfulness being recommended. Neither of these are objections as such, although they may point to some practical considerations for those concerned to promote playfulness in virtue epistemology.

A first set of issues concerns the range of epistemic virtues to which we are sensitive. Our inherited table of the virtues is deeply contingent, shaped by the vagaries of our cultural and intellectual history. Moreover, our *taste* for certain virtues is similarly structured by many biases and preferences, especially for what we might call “Yay!” virtues—those, like creativity, boldness and imaginativeness that are experienced, when exercised or observed, as *exciting*. By contrast, most people are less enthused about the “Yawn!” virtues—those, like diligence, carefulness and thoroughness that are typically dull, even if they are necessary (and, tellingly, especially necessary for tasks everyone regards as boring, like double-checking references or data entered into a spreadsheet). This is utterly natural. Playfulness connotes enjoyment and pleasure and excitement and respite from the obligations, work, and discipline of school-work or the office or the realities of daily life in an increasingly earnest, buttoned-down world.

The virtue of epistemic playfulness is obviously among the Yay! Virtues and that's no bad thing—but then one worries that people will be super-keen to cultivate and exercise that virtue to the neglect of the dull and unexciting Yawn! virtues. Everyone wants to play a game, but no-one wants to tidy it neatly back into its box. The practical question is then to ask how we can promote the virtue of epistemic playfulness while also instilling due regard for the Yawn! virtues—or what we might more formally call the procedural epistemic virtues. When tackling that question, the obvious people to consult are teachers and parents who are tasked with constantly trying to ensure equal appreciation of playfulness and seriousness and the arrays of Yay! and Yawn! virtues they represent.

A second set of issues concerns the relationship of epistemic playfulness to the wider normative commitments of the epistemic agent.

Obviously, all virtues are located in a wider constellation of values and goals and some conception, however inchoate, of the good life—think of a *eudaimone* Aristotelian, a consummate Confucian, or a “wholesome” Buddhist. If we run with that thought, we can ask about the emplacement of playfulness within wider philosophical visions of human life and the wider order of things.

If that sounds too high-falutin’, consider the example of early Daoist philosophy as exemplified by Zhuāngzǐ, the model *par excellence* of a certain form of epistemic playfulness. In his eponymous book, there is playfulness galore—absurd stories about gigantic birds, affectionate banter with a friend, weird stories designed to prick human conceits and lots of teasing criticisms of stuffy Confucians. A Zhuāngzǐst Daoist enjoyably manifests the virtue of epistemic playfulness for precisely the reasons given by Nguyen—dissolving dispositions to dogmatism and doxastic ossification, engaging in the determinedly perspective-shifting that Daoists call *yóu* (“wandering, roaming”) and constantly using humor and irony to expose and undercut what Nguyen calls “epistemic traps”. Crucially, the intelligibility and salience of this epistemic playfulness are provided by the wider vision of human life articulated by Zhuāngzǐ—the perspectival character of our specific “ways” of experiencing and engaging with the world, for instance, and the conviction that rigid and inflexible styles of action are not consonant with *Dào*. The playfulness reflects a properly enlightened relationship to the realities of human existence of the sort attained by the *zhēnrén*, the “true” or “authentic” person who “fathoms the real character of life” (*Zhuāngzǐ* ch. 19).

Granted, this is portentous stuff, but what Nguyen offers is a compelling account of the seriousness of playfulness. It captures something vital about the character of an epistemic practice that could honor the fun Yay! virtues as well as the dull Yawn! ones. It also suggests much deeper ways that playfulness can transform our life and conduct for the better—to a better “way”, as Zhuāngzǐ would say, of conducting our epistemic affairs.

References

- Molière, *The Misanthrope, Tartuffe, and Other Plays*, trans. Maya Slater (Oxford: Oxford University Press, 2001), xiv.
- Zhuangzi, *The Essential Writings with Selections from Traditional Commentaries*, trans. Brook Ziporyn (Indianapolis: Hackett, 2009).

9c Commentary from Lani Watson

Playfulness Versus Epistemic Traps

The question at the heart of Nguyen's chapter is both intriguing and unusual: is there "some kind of genuine and deep connection between playfulness and intellectual virtue" (p. 1)? As Nguyen puts it, "Some sages are full of humor, and some of the best insights start as jokes" (p. 1). It appears at least plausible that such a connection exists and is worth exploring. Nguyen provides an engaging exploration of this connection, presenting a challenge to the idea that all intellectually virtuous conduct must be guided by the serious and somber pursuit of truth.

This contention strikes me as essentially correct and the chapter offers good reasons for attending to the relationship between intellectual virtue and playfulness more generally. Beyond this, Nguyen argues that intellectual playfulness "is the right disposition to get us out of a certain kind of dogmatism" (p. 1). This argument raises further questions about the work that intellectual playfulness can do in the context of what Nguyen terms "epistemic traps", particularly those he calls "inquiry traps", paradigmatically illustrated by the case of echo chambers. This idea, I think, faces some challenges worth considering.

To begin, it is illuminating to examine the relationship between playfulness and intellectual virtue, as Nguyen presents it. He refers to intellectual playfulness as a "specific cognitive varietal" (p. 1) of playfulness and states that "Intellectual playfulness, loosely, is the disposition to try out new ideas, perspectives and systems of thought for the sheer joy of it" (p. 1). The focus on intellectual playfulness is necessary, given the scope and aims of the chapter. Nonetheless, this leaves the relationship between playfulness and intellectual playfulness mostly unexplored. In particular, it raises the question of what makes intellectual playfulness, distinctively intellectual and, in turn, the questions of how and whether intellectual playfulness is rightly classed an intellectual virtue.

Equivalent questions can, no doubt, be raised in the case of many, if not all, of the intellectual virtues. To my mind, however, they are particularly intriguing in the case of intellectual playfulness. For the most

part, when one sees “intellectual X” in the literature, the standalone “X” is a moral virtue, such as humility, perseverance, or courage, and the word “intellectual” is used to distinguish between this moral virtue and its intellectual counterpart. One can always probe further and ask what the word “intellectual” actually designates and there is, I think, no convincing consensus regarding this. However, at least one prominent response suggests that “intellectual” signifies the pursuit of distinctively intellectual ends; “a desire for the truth, for getting things right” as Croce and Pritchard put it (this volume). In other words, intellectual virtues are *essentially* concerned with the pursuit of intellectual ends.

Intellectual playfulness, as Nguyen defines it, problematizes this. As he emphasizes, play is autotelic; in other words, it is “done for its own sake” (p. 6). As such, intellectual playfulness is not defined in terms of intellectual ends. If that’s right, this raises a dilemma: either intellectual virtues are *not* essentially concerned with intellectual ends (and intellectual playfulness is indeed an intellectual virtue) or intellectual playfulness is not an intellectual virtue (because intellectual virtues *are* essentially concerned with intellectual ends).

I raise this dilemma, not because it presents a deep challenge to Nguyen’s central argument concerning the role and value of intellectual playfulness in our lives. That argument can be made without committing to the claim that intellectual playfulness is a virtue and perhaps it doesn’t much matter, in and of itself, whether we class intellectual playfulness as a virtue or not. Regardless, as I see it, this dilemma provides an opportunity to further define the nature of intellectual virtue itself. Investigation into what it is that makes intellectual playfulness distinctively intellectual, and in turn, a candidate for intellectual virtue, seems like promising grist for the mill. Nguyen’s exploration lays valuable groundwork for this.

Moreover, it is instructive to examine the relationship between intellectual playfulness and closely related intellectual virtues, such as curiosity and open-mindedness. Interestingly, Nguyen does not examine curiosity in the chapter, although one might think it stands out as an often, perhaps paradigmatically, playful intellectual disposition. Nguyen does, however, contrast intellectual playfulness with open-mindedness, focusing specifically on an essential component of intellectual playfulness, namely, perspective-shifting. He states; “Open-mindedness...turns out to be quite different from perspective-shifting” (p. 10) (although, somewhat confusingly concedes in a footnote that perspective-shifting might actually be a form of open-mindedness (strong open-mindedness; p. 11)).

The relationship between intellectual playfulness and open-mindedness is especially salient because of the emphasis Nguyen places on this contrast with respect to the central contention that intellectual

playfulness (not open-mindedness) is “the right disposition to get us out of a certain kind of dogmatism” (p. 1), exhibited in inquiry traps, such as echo chambers. Nguyen states:

Mere open-mindedness leads us to inquiries conducted while using our standing belief-system. And in an inquiry trap, that belief system has been rigged to re-affirm itself. If the function of open-mindedness is to iron out incoherencies in one’s belief system, then it won’t help against a trap belief system which has already been engineered for appealing internal consistency.

(pp. 9–10)

To my mind, Nguyen is almost certainly too quick to dismiss the capacity of open-mindedness to free a person from an inquiry trap. It seems like truly virtuous open-mindedness must have this capacity, indeed should be to some extent defined by it. At any rate, Nguyen argues that open-mindedness does not involve the kind of perspective-shifting that is required to escape an inquiry trap. Rather, this kind of perspective-shifting is found in intellectual playfulness because it, unlike open-mindedness, is autotelic.

This distinction represents the broader claim underpinning Nguyen’s central contention. It is the autotelic nature of intellectual playfulness that frees a person from what he calls the “plausibility constraints” of their standing belief system. Such plausibility constraints restrict the open-minded person, for example, by narrowing their focus to only those lines of inquiry that already appear plausible, based on their current beliefs. Given that intellectual playfulness does not have intellectual ends, it also does not, according to Nguyen, come with plausibility constraints. Hence, the intellectually playful person can escape an inquiry trap, such as an echo chamber, by happily shifting to a perspective (outside of the chamber), even when it doesn’t appear plausible to them. Perspective-shifting under non-autotelic conditions (such as in open-mindedness) fails on this score; “Even if you are trying out alternative systems of belief, the choice of those systems will still be influenced by your standing system of beliefs” (p. 12).

I think this argument requires further scrutiny. Intellectual playfulness cannot, after all, be an epistemically neutral activity. What we play and how we play are influenced by what we believe. Play is perhaps less likely to be affected by plausibility constraints (although it seems possible that plausibility is an important element of some forms of play), but it is surely no less shaped by other constraints that are, fundamentally, a product of the player’s belief. Even if one shifts perspectives for purely autotelic reasons, one is still constrained by a range of background

biasing factors, including, for example, the perceived palatability of the perspectives one is playing with. We might call this a kind of palatability constraint.

One can see the effects of palatability constraints by drawing on the example Nguyen uses to illustrate plausibility constraints. Nguyen contends that even someone virtuously motivated to pursue moral truths would be unwilling to take up the Nazi perspective because it is, among other things, implausible as a candidate for moral truth. A similar issue, however, seems to apply to the person engaged in intellectual playfulness. The intellectually playful person is, I think, no more likely to adopt the perspective of a Nazi, than the person seeking moral truths. In fact, I would hazard that they are less likely to shift into this perspective—because where would the “sheer joy and fun” be in that? The palatability constraints that come with intellectual playfulness are arguably more, not less, restrictive when it comes to contrasting ideologies and moral perspectives.

Nguyen describes the intellectually playful person as one “trying out and exploring systems of belief because they are funny, beautiful, elegant, or charmingly bizarre” (p. 13). It is notable that this list does not include trying out and exploring systems of belief because they are, for example, dangerous, undermining, unpleasant or threatening. Indeed, it seems plausible that intellectual playfulness might actively preclude the possibility of such exploration. This is a problem for the claim that intellectual playfulness can free us from inquiry traps, such as echo chambers. As Nguyen rightly points out, echo chambers function precisely to make other perspectives appear dangerous and threatening. As such, the intellectually playful person in an echo chamber-induced inquiry trap is, I think, still highly unlikely to escape, given the palatability constraints that are almost certainly in play in that setting.

Nguyen recognizes the limitations of intellectual playfulness, stating that we “need a diverse portfolio of perspective-shifting dispositions” (p. 14) and conceding that “playfulness won’t get us to intellectual virtue by itself” (p. 14). He nonetheless contends that intellectual playfulness, when adopted, can “drive us out of our usual intellectual paths, and encourage us to occasionally leap into faraway perspective” (p. 17). There is little to argue with in that and, as I said at the outset, Nguyen makes a good case for attending to the relationship between playfulness and intellectual virtue. There are, however, I think good reasons for being more cautious with respect to the stronger claim that intellectual playfulness can function as (even a partial) antidote to the pernicious and ever-pressing phenomenon of inquiry traps, such as echo chambers.

9d C. Thi Nguyen's Response to Commentaries

Rejoinder to Watson and Kidd

Let me begin by thanking Lani Watson and Ian James Kidd for their excellent, thoughtful, and delicate commentaries. In this brief response, I'd like to concentrate on only two of the issues raised by these commentaries—two that I find particularly interesting.

First, Watson wonders whether intellectual playfulness is really an intellectual virtue:

As [Nguyen] emphasises, play is autotelic; in other words, it is “done for its own sake” (p. 6). As such, intellectual playfulness is not defined in terms of intellectual ends. If that's right, this raises a dilemma: either intellectual virtues are not essentially concerned with intellectual ends (and intellectual playfulness is indeed an intellectual virtue) or intellectual playfulness is not an intellectual virtue (because intellectual virtues are essentially concerned with intellectual ends).

This comment cuts right to the issue of what's really fascinating about intellectual playfulness. The core idea here is that thinking about intellectual playfulness reveals that some epistemic ends are *self-effacing*. Self-effacing ends are those ends that aren't best achieved through direct pursuit. To get to the end, you have to aim at something else. As many have suggested: happiness might be a self-effacing end. You don't achieve happiness by pursuing happiness directly, but by absorbing yourself in the pursuit of something else: like knowledge, helping others, or success. And games, as I have suggested elsewhere, are good engines for pursuing self-effacing ends. The very structure of a game reveals a curious motivational inversion: in games, there is a local goal that we pursue, and a larger purpose that we play for. But often that larger purpose is not—and, in some cases, cannot—be at the forefront of our minds. I climb to relax, but in order to relax, I cannot directly pursue relaxation during the climb. I pursue relaxation precisely by putting that purpose out of my mind, and absorbing myself just in the pursuit of getting to the top.¹

Thinking about intellectual playfulness, for me, reveals a presumption within some accounts of the intellectual virtues—a presumption is at least gestured to within Watson's discussion. That presumption is that what it is to have an intellectual virtue is to be actively concerned with achieving that virtue's good. For example: to be epistemically virtuous, one might think, one must be *actively and consciously* concerned with the various epistemic goods—like truth, reliability, etc.

One horn of Watson's dilemma is that "intellectual virtues are not essentially concerned with intellectual ends". But what is it to be "essentially concerned" with intellectual ends? Does that mean that the *activity* is formulated such as to bring about those ends, or need the *agent* also be actively and consciously in pursuit of those ends? My suggestion is simply that intellectual playfulness is a virtue in the sense that will lead the possessor to better epistemic states—but that, in possessing and enacting it, the possessor will not be actively seeking better epistemic states. That is, the *activity* might be concerned with intellectual ends, but the undertaker might not be, in order to achieve those ends. In other words, the epistemically best-off agent will sometimes undertake intellectual activity for nonintellectual ends.

My background worry here is that some approaches to intellectual virtue theory might have *presumptively ruled out* the possibility that some of the desired goods here might have a self-effacing structure. I suggest that we deny that presumption, and make room for modes of virtue where certain character traits consciously aim an agent at one good, while actually achieving another. Like, for example, intellectual playfulness. (One might, instead, take the other fork and hold onto the view that to have a virtue, one had to actively pursue its associated good. But in that case, if you accept my analysis, we are led to the puzzling position of thinking the most intellectually virtuous agent is not the epistemically best-off agent.)

Onto the next concern. Both Watson and Kidd worry, in different ways, about the limitations of intellectual playfulness. Kidd's worry is that intellectual playfulness bends us towards the virtues associated with delight, but not toward the virtues that are boring and tiresome to enact. Watson's worry is similar: that in trading normal intellectual life for intellectual playfulness, we have traded *plausibility constraints* for *palatability constraints*. That is, in normal intellectual life we explore only those perspectives that seem to us plausible; but in intellectual playfulness, we explore only those perspectives that seem to us fun or delightful.

I entirely agree with these worries: this is an essential limitation of intellectual playfulness. But the point was never that intellectual playfulness could do it all. The point was that such playfulness was part of a balanced diet of intellectual virtues, each of which had its strengths and its constraints. Specifically, intellectual playfulness is supposed to be

something of a hedge and an insurance policy against a specific kind of epistemic trap. Other virtues are needed to balance out its weaknesses. It will not be conducive to epistemic virtue on its own; it's part of a package deal.

Another way to put it: each mode of intellectual pursuit arises from a motivation. It seems plausible to think that each motivation comes with certain constraints. Normal intellectual life is vulnerable to epistemic traps that modify plausibility. Intellectual playfulness is limited in its preference for what is delightful and fun. Another motivation for exploring perspectives might be empathy, which leads us to take on the perspective of people, not because it was fun or plausible, but because we cared about a particular person. But this comes with another constraint: it is constrained by the kinds of people we care about, and typically encounter (Bailey 2020).

The larger view that this suggests might be, to some, startling: there is no singular motivational state which it is always good to occupy. The epistemically virtuous might need is to be able to shift between motivations, or to act in concert with people with other motivations. Normal intellectual interest in epistemic goods, intellectual playfulness, empathy—each has its strengths, and its gaps. None is complete on its own; each requires assistance from other modes, to help move us towards covering possible gaps.

Note

- 1 For a further discussion, see (Nguyen 2019, 2020.)

References

- Bailey, Olivia. 2020. "Empathy and the value of humane understanding." *Philosophy and Phenomenological Research*. <https://doi.org/10.1111/phpr.12744>
- Nguyen, C. Thi. 2019. "Games and the art of agency." *Philosophical Review*. 128 (4): 423–462.
- . 2020. *Games: Agency as Art*. Oxford: Oxford University Press.

Part III

Collective Virtues
and Vices

T&F Proofs – Not for Distribution

T&F Proofs – Not for Distribution

10 Solidarity

Virtue or Vice?

Heather Battaly

In their 2016 analysis of “Collective Virtue,” Ryan Byerly and Meghan Byerly argue that some virtues are distinctively collective and suggest that solidarity is such a virtue. As they see it, a *distinctively* collective virtue is a virtue of a collective (or group) for which there is no individual analog; that is, there is no corresponding virtue V of individuals, from which the collective version of V could be derived. In proposing the virtue of solidarity as a paradigm case, Byerly and Byerly recognize that: “an account of collective solidarity cannot be derived from an account of individual solidarity... since there simply is no such thing as individual solidarity” (2016, 49). They rightly point out that: “an individual has no members that can empathize with and unite themselves to each other” (2016, 49). Sally Scholz likewise confirms, in *Political Solidarity*, that: “one cannot be in solidarity with oneself” (2008, 19). Here, I use Byerly and Byerly’s suggestion as a springboard for exploring the virtue of solidarity, and thereby hope to contribute to the broader project of examining a virtue that is distinctively collective. This chapter is exploratory in spirit. It brings virtue theory to bear on some key accounts of political solidarity, flagging several points of controversy along the way. My hope is that shining a spotlight on the virtue of solidarity will contribute to discussions at the intersection of social epistemology, virtue and vice epistemology, and political philosophy.

Any analysis of solidarity will need to account for a simple datum: groups of “bad actors” can have solidarity. Nazis can have solidarity, as can Mexican drug cartels, Russian troll factories, and American tobacco-industry executives. What this shows is that solidarity won’t always be a virtue. It won’t be a virtue in groups whose aims are bad—morally bad in the case of Nazis and drug cartels, epistemically bad in the case of troll factories and tobacco-industry executives (whose aim is to sow doubt). But, even when a group’s aims are good ones, solidarity won’t always be a virtue. Too much solidarity can be a bad thing for a group. For group members, it can result in uncritical deference, mindless outsourcing, and thoughtless conformity. For marginalized subgroups, it can result in self-silencing and self-censorship. All of which can culminate in the epistemic stagnation of the group as a whole. In short, too much solidarity can result in epistemic vices.

Drawing on Tommie Shelby's account of political solidarity in *We Who Are Dark* (2005), Section 1 proposes a working analysis of the *trait* of solidarity. It argues that a group's solidarity consists in a quintet of its member's dispositions. Roughly, a group has solidarity to the extent that its members are disposed to (1) share values, aims, or goals; (2) care about those values, aims, or goals; (3) act in accordance with those values, aims, or goals; (4) trust the testimony of other group members with respect to those values, aims, and goals; and (5) feel a sense of belonging to the group. In this manner, group solidarity is a "degree concept" and a "threshold concept." To have solidarity, enough of a group's members must meet the basic threshold of having the aforementioned dispositions. But, once this basic threshold is met, it can be exceeded to different degrees, depending on the relative strength of those dispositions.¹ This analysis of the *trait* of solidarity does not presuppose that the trait is always a *virtue*. On the contrary, Section 2 suggests that excesses of the trait of solidarity can result in (or constitute) vices, as can deficiencies of the trait. It proposes several such excesses, with a focus on those that are epistemic, including uncritical deference, the mindless outsourcing of one's beliefs to the group, the self-silencing of subgroups, and the resulting epistemic stagnation of the group as a whole. It further argues that for the trait of solidarity to be a *virtue*, the group must exercise good judgment, which reins in these excesses, in addition to having good aims and motives. In short, it uses the framework of virtue theory to explain when and why solidarity is good and what makes it so. The conclusion returns to the topic of whether solidarity is a distinctively collective virtue.

1 What Is the Trait of Solidarity?

Let's begin with Shelby's influential analysis of solidarity. Shelby proposes a set of five conditions that are necessary and jointly sufficient for, what he calls, "robust" solidarity—a form of solidarity that is "strong enough to move people to collective action" (Shelby 2002, 237). On his view, a group has the trait of robust solidarity to the extent that its members: (i) identify with one another or the group; (ii) are motivated by a special concern for one another or the group; (iii) share values or goals; (iv) are loyal to one another and the group; and (v) trust one another and the group (Shelby 2005, 68–70). Shelby famously uses this analysis of the trait of solidarity to develop a more specific notion of black political solidarity, the aim of which—eliminating anti-black racism—is good. But, he does not restrict robust solidarity to groups with good aims; he points out that, for example, political parties, militias, and crime syndicates can have such solidarity, even if their aims are bad. The key point, for present purposes, is that Shelby is analyzing the *trait* of solidarity without presupposing that this trait is good or virtuous.

Let's unpack his analysis. Shelby's first necessary condition is the *identification* of a group's members with one another or the group. On his view, such identification involves a feeling of belonging to the group, whereby group members "think...in terms of 'we' rather than 'I'" (2002, 238). It also often involves an empathic understanding of other group members such that one member feels pride when another does well and embarrassment when another does poorly, "almost as if one had done the deed oneself" (2005, 68). Shelby describes the second condition, *special concern*, as a motivation to help, assist, and comfort other group members (2005, 68). This concern is "special" in the sense that it is "partial" to group members (and distinct from an impartial moral duty to help others). He illustrates the notion as follows:

The members of a sport team that has solidarity will show special concern toward one another, so that when a member is injured or is not doing well otherwise, other members will offer comfort and support, even when this has no direct bearing on the team's collective goals.

(2005, 70)

His third condition requires members to *share some values or goals* and know or confidently believe that fellow members are likewise committed to those values or goals (2005, 69). Importantly, it allows these values and goals to be vague. To illustrate, Shelby argues that members of black political solidarity groups share a basic goal of eliminating anti-black racism that allows for disagreement over interpretations of that shared goal and over which actions, policies, and priorities will contribute to it (2005, 125, 247). Fourth, he argues that solidarity requires group members to be *loyal* to the group's goals and to other members, where this involves acting (and even exerting "extra effort") to help group members and to advance the group's goals (2005, 69). He notes that loyalty also involves sacrificing one's own interests and goals when they compete with those of the group. Finally, group members must *trust* one another and the group, especially since loyalty to the group sometimes requires members to make individual sacrifices to advance the group's interests. For Shelby, trust involves believing that other group members won't let one down or betray the values of the group.

While Shelby's analysis of solidarity offers five key insights, which will motivate the analysis I propose below, it also risks being too demanding. One worry is that special concern may be too demanding to be required for solidarity and may instead be necessary for a different concept, perhaps something in the neighborhood of caring, friendship, or community. To explain, special concern will be too demanding if it requires group members to care about the interests and goals of other members *even when* their interests and goals are unconnected to those of

the group.² Case in point, the American Association of University Professors can have solidarity in resisting a furlough even when its members don't care about one another's personal athletic goals. Indeed, a group's members need not even like one another very much to have solidarity.³ Tobacco-industry executives didn't need to like one another in order to have solidarity in lying about the risks of smoking. Nor need a group's members care about one another's general well-being to have solidarity.⁴ A drug cartel can have solidarity in making money and gaining power even if its members don't care about one another's well-being, or anyone else's. Analogous worries arise for the loyalty and trust conditions above. Solidarity doesn't require helping other members achieve goals, or trusting them to be faithful to goals, when those goals are unconnected to the goals of the group.⁵ In sum, this worry is about the *scope* of the motivations, actions, and trust needed for solidarity.

Further, empathy's role in solidarity may be exaggerated in the analyses of both Shelby and Byerly and Byerly. If empathy requires effortful perspective taking, and/or sharing and understanding other's emotions, it will be quite demanding and difficult to achieve.⁶ In which case, it isn't likely to play the role that Shelby has in mind; that is, it isn't likely to be a default route to identifying with a group, or feeling "the familiar sense of 'we-ness' that is...characteristic of solidarity groups" (Shelby 2005, 68). To illustrate, one can identify with a political party, feel a sense of belonging to it, and think of its victories as one's own (e.g., "We won the congressional seat back!"), even if empathy with other members of the party is neither present nor in the offing. Perhaps, we can expect empathy, when present, to increase one's sense of belonging, but the point is that we can't assume empathy to be present.

The worries above point us toward a less demanding analysis of solidarity that still preserves five of Shelby's insights. The **first insight** is that solidarity requires group members to:

- (1) share some values, interests, aims, or goals.

As Shelby is at pains to emphasize, group members need not share *all* of their values or goals. What they must share is *some* basic value, interest, or goal in broad outline. If they fall short of sharing even that much, then they won't meet the minimum threshold for solidarity. Accordingly, to satisfy this condition, members of Black Lives Matter need only share the basic goal of "working for a world where Black lives are no longer systematically targeted for demise."⁷ They need not share views about how best to interpret or achieve that goal, or priorities with respect to subgoals. In this manner, Black Lives Matter (BLM) can meet the minimum threshold for solidarity even when different group members—for example, political liberals, feminist progressives, Marxists, religious conservatives—prioritize different subgoals. Still, solidarity is also a

degree concept. This means that even if a group can meet the minimum threshold for solidarity by sharing a very basic goal, groups whose members also share many of their subgoals will have more solidarity than groups whose members share fewer subgoals or only very basic goals (other things being equal).⁸

Let's now home in on what it means for group members to *share* some basic values or goals. Here, I employ a thin sense of sharing, whereby two or more individuals count as sharing values or goals whenever they in fact have the same evaluative beliefs and commitments—whenever their evaluative beliefs and commitments are in de facto agreement. Accordingly, all individuals who believe that it is good to reduce carbon emissions thereby share a basic value, and all individuals who are evaluatively committed to reducing carbon emissions thereby share a basic goal. Nothing more is required for sharing a basic value or goal.⁹ In particular, these individuals need not be motivated to reduce carbon emissions to value their reduction, nor need they care very much about reducing them to have the goal of reducing them—their other values and goals (which prioritize convenience) may often or always defeat their goal to reduce them. Nor need they often or ever act so as to reduce carbon emissions; they can have and share values and goals that they consistently betray.

This brings us to the **second insight**: solidarity requires some sort of consistent motivation or care on the part of group members. Perhaps the following is an obvious point: it won't be enough to merely share values and goals in the thin sense described above, if group members aren't consistently motivated to pursue them or don't care much about them. To put this differently, if individuals can share values and goals without caring, or caring much, about them, and without acting in pursuit of them, then it is hardly surprising that sharing values and goals is insufficient for solidarity. Granted, on the view proposed here, failing to share any values or goals will be *one* way of falling short of the trait of solidarity. But, importantly, it won't be the only way—group members can also fall short by consistently betraying shared values and goals. To avoid such betrayals, they will at least need to consistently care about and be motivated to pursue shared values. Note that ephemeral, “one-off” motivations to pursue their shared values won't be enough: at a minimum, they will need consistent motivations to pursue shared values—motivational dispositions.¹⁰ Accordingly, I propose that solidarity at least requires them to:

- (2) be disposed to care about acting in accordance with their shared values, interests, aims, or goals.

Condition (2) explicitly supplies a disposition to care about and be motivated to act in accordance with values, goals, and the like, while

simultaneously restricting the scope of that disposition to values (etc.) which are *shared*. It thus obviates any need to appeal to special concern as the source of motivation and likewise avoids blurring the line between solidarity, caring, and friendship. Importantly, theorists who employ thicker interpretations of values themselves, or of sharing values, including theorists who think such sharing involves “joint commitment,” may assume that (1) entails (2). Theorists may likewise assume that having a goal entails having some motivation to pursue it—however weak and defeasible that motivation may be—in which case, (1) would entail (2) as far as goals are concerned. I have assumed this much about goals, but not about values; I think this marks a difference between goals and values. We’ll return to goals below. Let’s now anticipate two broader questions. First, contra (1), won’t there be cases of solidarity without any shared values, goals(etc.)? Lawrence Blum (2007) raises this objection, arguing that solidarity does not conceptually require sharing a basic value, interest, or goal. On his view, sharing values (etc.) *can* be a causal basis for solidarity, but so can sharing one’s “thin” identity (e.g., being black), or sharing one’s experiences (e.g., of anti-black racism), in the absence of any shared values, interests, or goals (2007, 63).¹¹ In short, he argues that sharing one’s “thin” identity and/or experiences with others can ground a sense of belonging, mutual concern, and mutual support—features which he takes to be key for solidarity. Let’s assume for the moment that sharing a “thin” identity, or sharing experiences, can causally lead to a sense of belonging and mutual concern and support. The problem is that like Shelby’s special concern condition, the features Blum mentions—especially, mutual concern and support—seem to target friendship or community, which aren’t required for solidarity.¹² So, even if shared identity or experiences can be a causal basis for mutual concern and support, that doesn’t give us reason to think that shared identity or experiences—in the absence of shared values—can be a causal basis for solidarity. As argued above, we can avoid targeting friendship and community by restricting the scope of the motivational component in (2) to values, interests, or goals that are shared. Further, we can even acknowledge that shared identity and/or experiences sometimes causally lead to solidarity rather than to friendship or community. But, importantly, these will be cases in which the shared identity or experience has causally led to solidarity because it has led to some shared values, interests, or goals. These values (etc.) need only be basic ones, as when a shared experience of racism generates a shared evaluative belief that racism is bad, or when the shared identity of being a woman generates a shared evaluative commitment to “girl power.” Relatedly, we can acknowledge that having a shared fate sometimes leads to solidarity, but these will also be cases in which the shared fate leads to shared values, interests, or goals.

Second, the above argued that for group members to share values and goals in the thin sense, they must at least have the same evaluative beliefs and commitments. Does solidarity also require them to *know* that fellow members have the same values, interests, or goals, and be “jointly committed” to those values, interests, or goals?¹³ Shelby (2005) argues that it does, Scholz (2008) argues that it doesn’t, and Adam Cureton (2012) takes the middle road. Shelby’s answer allows for solidarity among smaller groups in which members know one another and make decisions together, for example, local chapters of labor unions. But, it risks preventing larger groups, whose members are spread out (across the globe or across time), from having solidarity. It arguably precludes global NGO’s from having solidarity, since their members won’t usually know one another’s level of commitment. In contrast, Scholz’s answer allows for solidarity among larger groups, but at the potential cost of counting any random collection of individuals, who de facto have the same goals (and de facto meet her other conditions), as members of a solidarity group (Scholz 2008, 56, 115–116, 121–122). It risks going too far in the other direction. Cureton tries to avoid both of these risks. On his view, members of a solidarity group need not know one another, nor need they know the details of one another’s commitment levels or the specifics of the projects they have in common. But, they must know that “they are working with others and know in broad outline the nature of the cooperative activities in which they are taking part” (2012, 699). I am hopeful about the prospects for Cureton’s middle road, as it has the potential to extend solidarity to larger groups, without extending it to random collections of individuals.¹⁴ Nevertheless, I flag this as an issue that warrants further exploration and defense, since a confident answer would need to engage with Margaret Gilbert’s (2014) work on joint commitment and collective goals.

Note that whatever answer we end up giving, (1) and (2) will still be insufficient for solidarity because they still allow for consistent betrayal. For a group to avoid consistent betrayal and have solidarity, its members will need more than shared goals and dispositions to care about them; they will also need dispositions to *act* in accordance with those goals.¹⁵ Perhaps, this point is also obvious. But, to explain: group members can share goals and share *motivational dispositions* to act in accordance with them, while still failing to be *disposed to act* in accordance with them because they have even stronger motivational dispositions to do something else, such as, what is easy or expedient.¹⁶ As alluded to above, many of us have motivational dispositions to reduce carbon emissions, but also have even stronger (and defeating) motivational dispositions to do what is easy. Similarly, group members may fail to be disposed to act in accordance with shared goals whenever there are burdens of doing so and effort is required, for example, whenever it comes time to do something more than merely click a link.¹⁷ In other words, they

may be disposed to betray those shared goals whenever it isn't easy or convenient to act in accordance with them. And, insofar as they are disposed to betray those shared goals at the first sign of inconvenience, their group will fail to meet the minimum threshold for solidarity. Their group may meet the threshold for something like "fair-weather" support, or "warm-glow" giving, but it will fall short of solidarity, which requires having a motivational disposition that is strong enough to generate a disposition to act in accordance with shared goals, even when doing so involves some sacrifice. To use Shelby's terminology, the trait of solidarity is "robust" (Shelby 2002, 237).

This leads us to Shelby's **third insight**: solidarity requires loyally acting in accordance with shared goals. As I'll put it, solidarity requires group members to be disposed to act in accordance with shared goals, even when doing so involves some effort or inconvenience. Scholz clarifies this point, arguing that the disposition to act in accordance with shared goals isn't "one size fits all." It can be manifested in different actions, in different group members, that draw on "each person's talents and abilities" (2007, 85). To illustrate, if the shared goal is eliminating anti-black racism, we can expect a wide range of actions to manifest this disposition, including donating funds, protesting, consciousness-raising, action planning, policy research and writing, and (for some) daily efforts to survive anti-black racism. Still, whichever actions group members perform in manifesting this disposition, the key point is that meeting the minimum threshold for solidarity will require them to be disposed to make some sacrifices of time and effort. Since solidarity is a degree concept, groups whose members are disposed to make many sacrifices, including putting off their own goals, will have more solidarity than groups whose members are disposed to make minimal or fewer sacrifices.

What exactly does it take to *act in accordance with* a shared goal? Must group members have detailed knowledge about the action plans of the group (or participate in decision-making about those plans as members of a local chapter of a labor union might), and perform the actions the plan has assigned them? If not, must they know about the actions of fellow members in broad outline and coordinate their own actions with these? Or, will it suffice to perform actions that de facto contribute to the goal, without any knowledge of what other members are doing? As above, the first answer risks precluding larger groups from having solidarity, whereas the third answer threatens to cast the net too widely. Indeed, there is good reason to reject the third answer. As Avery Kolers (2016, 52–53) has convincingly argued, a group might have condemned some of the actions that de facto contribute to the goal, in which case members who perform such actions would not be acting in solidarity with the group. To illustrate, if I correctly believe that I can reduce carbon emissions by recycling aluminum, but the environmental group to which I belong has recently rejected the practice of recycling (citing new

evidence that it doesn't succeed in reducing net emissions), then in recycling aluminum I do (we are assuming) reduce carbon emissions, but I don't act in solidarity with the group. To act in solidarity with the group, I must at least perform actions that they don't reject. That is a minimum requirement for coordinating my actions with theirs. Here, as above, I am hopeful about the advantages of the middle road, but flag this as an answer that also requires further defense. I likewise flag the related subject of joint and collective action.¹⁸ We will need to determine whether and how the actions of individual group members contribute to the joint or collective actions of the group. And, whether the dispositions of action that we are here requiring of individual members will need to be dispositions to act jointly or work together with other members toward shared goals. For now, we can summarize the third insight as follows. Solidarity requires group members to:

- (3) be disposed to act in accordance with shared values, interests, aims, or goals, even when doing so involves some sacrifice (of time, effort, or convenience).

Shelby's **fourth insight** is that solidarity requires trusting other group members. On his view, solidarity requires trust because solidarity requires loyalty, and loyalty requires trust: group members won't be loyal to a shared goal, and won't make the sacrifices loyalty entails, if they don't trust fellow members to make sacrifices of their own. But, arguably, loyalty *doesn't* require trust—a family member can be loyal to the family's shared goal of managing a jointly owned property even when she knows that the other members of her family will be too selfish to make any sacrifices.

Still, solidarity does require trust, even if loyalty doesn't. Why? As long as we take the middle road suggested above, acting in accordance with shared goals will entail knowing in broad outline that one is working with others who have the same goals, knowing roughly what kinds of actions they are performing in pursuit of those goals, and coordinating one's actions with theirs. But, first, we won't gain such *knowledge* if we don't trust the testimony of our fellow members when they report their goals and actions; knowing these things requires trusting their testimony about their goals and their actions. Accordingly, solidarity requires group members to:

- (4a) be disposed to trust the testimony of other group members with respect to their goals and actions.

Second, we won't be disposed to *coordinate our actions* with theirs, when their judgments about which actions to perform conflict with our own, if we never trust or defer to their judgments. Coordinating

our actions with theirs in cases of disagreement about which actions to perform requires a disposition to sometimes trust and defer to their judgments. Roughly, it requires a disposition: to sometimes change our minds about how to proceed and come to believe what they believe; or when that is not in the offing, to sometimes have more epistemic confidence in their beliefs than our own; or when neither is in the offing, to sometimes believe that it is more important (all things considered) to follow their judgments about how to proceed than our own.¹⁹ As Kolers puts it, “solidarity is deferential” (2012, 367). To see why, consider cases in which a member (M) of the National Coalition against Domestic Violence disagrees with the group about which actions will be effective; for example, M believes that intimate partner abuse would be reduced if victims confronted their abusers, whereas the majority of the group rejects this on the grounds that it puts victims at greater risk. If M never trusts other group members and never defers to their judgments, then M’s actions won’t (consistently) be coordinated with the group’s in these cases, and will even (consistently) undermine the group’s actions. In short, to be disposed to coordinate our actions with the group’s, when our judgments are initially in conflict with the group’s, we must be disposed to at least sometimes trust, and defer to, the group’s judgment. Of course, we might be disposed to be even more trusting than that—we might even be so trusting that we abstain from making any of our own judgments about which actions will be effective and rely entirely on the judgments of the group. But, while higher degrees of trust secure higher degrees of solidarity, they aren’t required for the disposition to coordinate our actions with the group’s, since enough of the group’s members—including those of the majority opinion—will also need to (be disposed to) sometimes trust and defer to members who disagree. Rather, all that is required for the disposition to coordinate our actions with the group’s (in cases of disagreement about which actions to perform), is the disposition to at least sometimes trust and defer to other members of the group. In other words, meeting the minimum threshold for solidarity requires group members to:

- (4b) be disposed to sometimes trust and defer to other members’ judgments about which actions to perform, even when they (at least initially) conflict with the group member’s own judgment.

As we will see in the next section, groups that manifest degrees of trust and deference that exceed this threshold (e.g., mindless trust and uncritical deference) will have more solidarity than groups whose dispositions of trust and deference merely meet it.²⁰

Finally, Shelby’s **fifth insight** is that solidarity requires a feeling of belonging to the group and a tendency to see oneself as part of the group

(e.g., “We won the congressional seat back!”). While I reject Shelby’s suggestion that this feeling of “we-ness” is connected to empathy, I propose that this feeling is still required for solidarity. If so, groups whose members tend to lack this feeling of belonging (perhaps because they instead feel alienated from the group) will fall short of the minimum threshold for solidarity.²¹ Accordingly, solidarity will require group members to:

(5) be disposed to feel a sense of belonging to the group.

While this requirement is the most tenuous of the five, retaining it may have an advantage: it may help us distinguish allyship from solidarity.²² Allies, arguably, satisfy conditions (1)–(4), even though they don’t feel a sense of belonging to the group. Accordingly, the proposal is that any individual (or group) who satisfies (1)–(4) but not (5) with respect to a target group G, whether or not they are members of G, would be an ally of G but wouldn’t be in solidarity with G. To illustrate, as an ally of the National Union of Mineworkers (NUM) in Britain, LGSM (Lesbians and Gays Support the Miners) protested and raised funds in support of NUM’s 1984 mining strike. Arguably, the members of LGSM satisfied conditions (1)–(4) above, but didn’t feel a sense of belonging to NUM.²³ On my view, this makes them allies of NUM, though they fall short of being in solidarity with NUM. Now, that may sound odd—we might think LGSM *was* in solidarity with NUM. But, in reply, conditions (1)–(4) go a long way in addressing those concerns. Crucially, we will still assert that LGSM (and its members) acted in solidarity with NUM, in the sense that LGSM (and its members) performed the same actions that a group in solidarity with NUM would perform. That comes with satisfying condition (3) above and is not a problem for the account.²⁴ LGSM likewise shared some aims with NUM, was motivated to pursue them, and trusted NUM with respect to those aims, as captured by conditions (1), (2), and (4). In addition, we can point out that even if members of LGSM did not feel a sense of belonging to NUM, they may well have felt a sense of belonging to the larger group composed of LGSM and NUM which developed over time, and that larger group of LGSM and NUM might itself have had solidarity. If those replies succeed, they deflate the force of the objection.

Let’s summarize the proposed analysis of solidarity as follows. A group has the *trait* of solidarity to the extent that its members are disposed to (1) share values, aims, or goals (in the thin sense described above); (2) care about those values, aims, or goals; (3) act in accordance with those values, aims, or goals; (4) trust the testimony of other group members with respect to those values, aims, and goals; and (5) feel a sense of belonging to the group.²⁵ While many of the above examples

have involved values and goals that are in some way political, that is not necessary. Sports teams can have the trait of solidarity (Shelby 2005, 70), as can (e.g.) families, ER departments, writing rooms for television shows, hiring committees, and research teams. My hope is that the above offers a working analysis of the trait of general solidarity. Political solidarity is one kind of general solidarity, which pertains to when the shared values and goals are political. Intellectual solidarity is another kind of general solidarity, which pertains when the shared values and goals are intellectual, as they are with research teams.

2 Is the trait of solidarity a virtue or a vice?

The above analysis of the *trait* of solidarity does not assume that solidarity is always a virtue. Nor should it, since there are clear cases in which it isn't: recall that white supremacists and tobacco executives can have solidarity. There are additional advantages to separating our analysis of solidarity as a trait, from our investigation into its status as a virtue or a vice. Doing so can help us home in on what makes the trait of solidarity a virtue when it is one. This general approach to traits and virtues is an example of what Ian James Kidd calls "normative contextualism": it initially conceives of traits as normatively neutral, and then investigates what turns those neutral traits into virtues or vices (Kidd 2020, 81).

This section proposes conditions on virtue that are Aristotelian in spirit. For the trait of solidarity to be a *virtue*: the group's shared aims (goals, etc.) must be good; enough group members must have good ulterior motives in pursuing those shared aims (etc.); and enough group members must exercise good judgment in their pursuit of these aims (etc.), including good judgment in making sacrifices, and in trusting fellow group members. The key point below is that good aims (etc.) and good motives (on the part of a group's members) won't be enough to make the group's trait of solidarity a virtue; group members will also need good judgment. Without good judgment, their dispositions (of action, trust, and so forth) can be excessive and result in vices—for instance, they can be disposed to make too many sacrifices, and be too trusting and deferential. Good judgment is needed to rein in these excesses.²⁶

For starters, let's address the **shared aims**, goals, and values in condition (1). We can note that the trait of solidarity will not be a virtue in groups whose shared aims are bad and whose values (evaluative beliefs) are erroneous. Solidarity will not be a virtue in a group of white supremacists—who aim to subordinate persons of color and falsely believe that whites are morally superior. Nor will it be a virtue in a troll factory—that aims to sow falsehoods and blithely dismisses the value of truth. Why won't it be a virtue in these groups? Let's briefly canvass

two Aristotelian answers. First, virtues require knowledge, rather than false evaluative beliefs (NE.1105a31). On Aristotle's view, virtuous people and vicious people all aim at what they *believe* is good, but virtuous people have knowledge of the good whereas vicious people do not (NE.1151a6–7). Second, virtues are valuable, and part of what makes them valuable are the valuable ends at which they aim. As Robert Adams puts the point, virtues are ways of “being *for* the good” (2006, Ch. 2). Thus, if a trait aims at ends that are not good, it won't be a virtue.²⁷ The main upshot is this: for the trait of solidarity to be a virtue, the shared aims must be good and the shared evaluative beliefs (values) must (at least) be true or constitute knowledge.

But, this won't be enough—the trait of solidarity won't be a virtue in groups whose aims are good, but whose members have bad **ulterior motives**. Consider a group that has the shared aim of reducing carbon emissions, and whose members satisfy conditions (1)–(5) above but are motivated to satisfy those conditions because they have invested heavily in renewable energy and stand to make billions. This group has a good aim (reducing emissions). Its members are also disposed to act in accordance with that aim and trust the testimony of fellow members with respect to that aim. Their proximate motive in condition (2) is a good one (insofar as it is good to “be *for*” aims that are good). But, their ulterior motive is bad—they are motivated by greed, and not by moral goods such as well-being or epistemic goods such as truth—which disqualifies their solidarity from being a virtue. On standard Aristotelian analyses, virtues require acting (etc.) for the right reasons and with the right motives. This group fails to meet that condition. As Nicolas Bommarito insightfully observes: “someone's attendance at a protest doesn't reflect well on their political engagement if they attend only because of their romantic interest in the protest organizer” (2016, 449).

Importantly, good aims and good motives still won't be enough—the trait of solidarity won't be a virtue in groups whose aims and motives are good, but whose **judgment** is poor. To explain, group members can have good aims and good motives, and have the dispositions in conditions (1)–(5) above, but if their judgment is poor, they can have those dispositions to excess. In other words, group members who lack good judgment can go overboard. They can be disposed to trust the testimony of their fellow members with respect to the good aims they share, but be *too trusting*. They can likewise be disposed to make sacrifices in acting to achieve those aims, but *make too many sacrifices*. To use Aristotelian language, they can be trusting and make sacrifices in some of the wrong circumstances, the wrong ways, and with respect to the wrong members. In short, they can lack *phronesis*. When they do, the trait of solidarity won't be a virtue in their group.²⁸

Let's begin to sketch a picture of what some of these excesses might look like, beginning with excesses of trust. Recall that condition (4b)

requires group members to be disposed to sometimes trust and defer to fellow members' judgments about which actions to perform, even when they conflict with their own. (4b) applies to judgments about which actions will be effective in attaining the shared goal, how differently situated members can best help, how to interpret the shared goal, and which subgoals to prioritize. Accordingly, we can assume that group members will be *too trusting* when their trust goes overboard. When they are so trusting that they trust, and defer to, the judgments of fellow members over their own, *even* in circumstances in which they shouldn't trust (or defer to) the judgments of fellow members; and, when they outsource their judgments to fellow members, and thus don't make any judgments of their own, *even* in circumstances in which they shouldn't outsource.

Of course, the tough question is: when should, and shouldn't, group members defer? While a confident answer to that question would require significant exploration, we can at least begin with the proposal that a group member should (only) defer to fellow members who have more knowledge and expertise than she does with respect to the actions and subgoals in question. To illustrate, it is presumably appropriate for a group member who aims to fight climate change to outsource the majority of her judgments to a knowledgeable subgroup of scholarly experts. But, crucially, not all of the knowledge relevant to a group's shared actions and sub-goals will come from *scholarly* experts. We can expect some of that knowledge to come from experience. As Kolers (2016, 90) points out:

In many cases where solidarity is invoked, those who invoke it are plainly in a better position than their audience to know details of the situation: it is workers in a particular factory who know the working conditions in that factory; it is victims of spousal abuse who know what sorts of measures tend to exacerbate the problem and which ones make it worse; it is people who have been to prison who are familiar with conditions there...

We can add that populations that are experiencing acute effects of climate change are in a better position to know the details of the effects of climate change on their environments, and persons of color are in a better position to know the details of the manifestations of racism in their lives.²⁹ Recall Laurence Thomas' prescient suggestion that deferring to the concerns and experiences of oppressed persons is appropriate in collective efforts to resist oppression. As Thomas (1993), and more recently Cherry (2017) have emphasized, group members who share the goal of resisting oppression can have different experiences of oppression, which makes it appropriate to "listen and give credence to

the testimony of others whose experiences are different from our own” (Cherry 2017).

With the above in mind, we can offer two (very) defeasible suggestions about being *too trusting*. First, it is inappropriate for a group member to defer to fellow members who have less knowledge and expertise than she does about the actions (etc.) in question. Second, in cases where group members are too trusting, they are disposed to defer, and/or outsource, to fellow group members, *whether or not* they have more knowledge and expertise about the actions (etc.) in question. In other words, they are too trusting when they uncritically defer or mindlessly outsource.

Obviously, this is merely a sketch and further exploration is required. For starters, we will need to address whether it is sometimes appropriate, *all things considered*, to act as a group member advises even though we know that the group member’s judgment is false, or that she is unreliable, or that her argument is misleading (Fantl 2018, 140). After all, moral values may sometimes trump epistemic values.³⁰ We will also need to explore exactly what makes uncritical deference and mindless outsourcing epistemic vices, and whether uncritical deference and mindless outsourcing are always epistemic vices. That will require a consult from vice epistemology.³¹

Let’s briefly turn to *making too many sacrifices*, where my sketch will be even more exploratory. Recall that condition (3) requires group members to be disposed to act in accordance with shared aims, even when doing so involves making *some* sacrifices. Which sacrifices are appropriate (with respect to which aims), and which are excessive? Here, it will be especially important to consider the sacrifices of members of subgroups, themselves marginalized within the group as a whole. For instance, it will be important to consider the sacrifices black women have made for the black liberation movement (Cherry 2020, 7). In Cherry’s poignant words: “Women showed up for the movement, but people rarely showed up for them” (2020, 7). As Cherry explains, black women have historically been targets of silencing within the movement—they have, to some extent, supported a movement that has tried to silence them.³² Have their sacrifices as a marginalized subgroup been appropriate, or excessive, with respect to condition (3)?

I hope it is obvious that this is a difficult question to answer! For starters, answering it would at least require consulting with political philosophers, political scientists, and political historians, who specialize in the black liberation movement, and with black women who are in the movement.³³ It should also be obvious that it will be difficult to determine how to “hit the Aristotelian mean” with respect to making sacrifices as a member of a marginalized subgroup.

Where does this leave us? Can we suggest any *clear* cases of excessive sacrifice on the part of a marginalized subgroup? We can at least suggest

the following: the dispositions of self-silencing and self-censorship are likely candidates for excessive sacrifice. Roughly, suppose that a marginalized subgroup is disposed to self-censor and self-silence to the extent that its members are disposed to censor and silence their own views about which actions to perform (etc.), in interactions with the larger group, *whenever* their views conflict with those of the larger group. Thus, members of the subgroup go along with the views of the larger group without raising their own dissenting views. Crucially, this will include censoring and silencing their own views, and going along with the views of the larger group, even when members of the subgroup have *more* knowledge and expertise than members of the larger group. Accordingly, we can (very) defeasibly suggest that a marginalized sub-group that is disposed to self-censor and self-silence in these ways is making too many sacrifices. (Of course, the subgroup need not be blameworthy for this.) Here, too, further exploration is required. At a minimum, we would need further analysis of self-censorship and self-silencing, of whether they are always epistemic vices, and of what makes them epistemic vices when they are. We would also need to consider whether such self-censorship and self-silencing are sometimes appropriate *all things considered*, for example, in order to survive.

Finally, we can expect excessive trust and excessive sacrifice to produce additional epistemic vices in group members, as well as epistemic vices in the group as a whole. With respect to group members, we can expect mindless outsourcing to result in thoughtless conformity, whereby group members share all the subgoals and judgments of fellow members, whether or not those fellow members are knowledgeable. More importantly, we can expect the aforementioned excesses of group members to result in epistemic vices in the group as a whole. In this vein, Shelby warns of “group-think,” closed-mindedness, and dogmatism in the group as a whole, arguing that they lead to “defective collective decision-making” (2005, 233).³⁴ We can add epistemic stagnation and wheel spinning to this list, when subgroup members with dissenting views self-silence, and other members of the larger group all uncritically defer to one another. Depending on the composition of the group as a whole and its leadership structure, the self-silencing of marginalized subgroups, when combined with the uncritical deference of other members to group leaders and the closed-mindedness of group leaders, can also lead the group as a whole to adopt progressively more extreme positions (i.e., to become polarized). As Olúfẹmi Táíwò (2020) puts a similar point about elites within a group capturing the group’s values: “In the absence of the right kind of checks or constraints, [elites] will capture the group’s values, forcing people to coordinate on a narrower social project than the group would if power were distributed differently.”

While this section has focused on *excesses* of trust and action that can result in (or constitute) vices, we can also expect *deficiencies* of

trust and action to result in (or constitute) vices. Betrayal is one such deficiency (mentioned in Section 1), as is distrust of testimony about actions and subgoals, some cases of which are testimonially unjust.³⁵ To illustrate the latter point, white feminists have (at least sometimes) been deficient in trusting the testimony of black feminists and LGBT feminists (Cherry 2020, 6; Scholz 2008, 140). Likewise, white members of antiracist groups have (at least sometimes) been deficient in trusting the testimony of black members (e.g., “The best way for you to help as a white member is consciousness-raising”). More broadly, we can explore deficiencies, and excesses, of any of the dispositions in (1)–(5), and examine what makes them vices, and determine whether they are always vices.³⁶

To sum up, for a group to have the trait of solidarity, enough of its members must have the dispositions captured in (1)–(5), which means they must not be deficient with respect to those dispositions. Moreover, for the trait of solidarity to be a virtue, the shared aims of the group must be good, enough of its members must have good ulterior motives, and enough of its members must exercise good judgment, which enables them to avoid excesses of dispositions (1)–(5).

3 Is Solidarity a Distinctively Collective Virtue?

This chapter has proposed an analysis of the trait of solidarity, argued that the trait is not always a virtue, and explored what would be required for the trait to be a virtue. It has been argued that a group will have the trait of solidarity to the extent that its members have dispositions to share aims, and make sacrifices, and trust fellow group members, in pursuit of these aims. It has likewise argued that the group’s trait of solidarity will be a virtue to the extent that its members share aims that are good, have good ulterior motives in pursuing these aims, and exercise good judgment in pursuing these aims. In other words, it has analyzed a *group’s* traits and virtue of solidarity in terms of the traits and virtues of its *individual members*.

Let’s close with two sets of questions about method that are well worth exploring. First, does this mean that solidarity is not a distinctively collective virtue after all? Recall that for Byerly and Byerly, a distinctively collective virtue is a virtue of a group for which there is no individual analog: there is no corresponding virtue V of individuals, from which the collective version of V could be derived. While the analyses above satisfy this condition—they don’t derive a group’s solidarity from the *solidarity* of its individual members—one cannot help but suspect cheating. Those analyses are, after all, “summative” in spirit, if not in letter (Fricker 2010). They still derive a group-level trait and virtue from the traits and virtues of its individual members. That is, they assume that when enough of a group’s members possess dispositions (1)–(5), the group

itself will possess the trait of solidarity, and (roughly) when enough of its members share good aims, have good ulterior motives, and exercise good judgment, the group itself will possess the virtue of solidarity. This means that *if* an individual member's possessing (1)–(5) amounts to their possessing (say) the trait of integrity, then our analyses would be deriving group-level solidarity from the integrity of its individual members. What does this show? If this is cheating, then it shows that we need further analysis of the connection between summativism and virtues that are distinctively collective. And, if it isn't cheating, then it shows that we need an account of why we are inclined to think it is.

Second, should we expect distinctively collective virtues, at the group level, to be associated with vices that are also distinctively collective at the group level? Why or why not? Relatedly, does the above sketch suggest that there are some vices associated with solidarity—for example, epistemic stagnation, closed-mindedness—that are *not* distinctively collective (in Byerly and Byerly's sense)? If so, are there other associated vices of excess that are distinctively collective—for example, group-think and group polarization (Broncano-Berrocal and Carter 2021)? Which, if any, vices associated with the virtue of solidarity will be distinctively collective?

Acknowledgments

I am grateful to Eric Berg, T.J. Broy, Charlie Crerar, Colin Klein, Kevin Maroney, Heather Muraviov, Katie Peters, Clifford Roth, and Katrina van Dyke for discussion, and to Mark Alfano, Ryan Byerly, Jeroen de Ridder, and Duncan Pritchard for comments.

Notes

- 1 On threshold concepts, see Swanton (2003, 63).
- 2 Cherry (2020) takes Shelby's notion of special concern to be restricted to caring about members' interests only when they are connected to the shared interests of the group.
- 3 See Blum (2007), which contrasts solidarity with community.
- 4 Cherry (2020) develops a distinct notion of 'solidarity care,' in which group members care about one another's well-being. She argues that solidarity care is not necessary for group solidarity: it is an important complement to, rather than a replacement of, group solidarity.
- 5 Relatedly, see Scholz (2008, 46, 48, 81, 117), which contrasts political solidarity with social solidarity, camaraderie, and friendship.
- 6 Coplan (2011) argues that empathy involves sharing affect while maintaining self-other differentiation. For competing analyses of empathy, see Coplan and Goldie (2011).
- 7 <https://blacklivesmatter.com/about/>
- 8 We can expect the former group to have more solidarity than the latter on the assumption that both groups also satisfy the other necessary conditions for solidarity.

- 9 Scholz (2008) seems to endorse this view.
- 10 The threshold can be exceeded to different degrees; i.e., among motivational dispositions that meet the threshold, some will have higher rates of consistency than others.
- 11 Shelby contrasts having a ‘thin’ black identity—e.g., fitting a particular phenotypic profile—with having a ‘thick’ black identity—e.g., participating in a culture that is distinctively black (2005, 209–212).
- 12 Blum argues that unlike community, solidarity is a response to threat or adversity. But, arguably, adversity isn’t necessary for solidarity.
- 13 On joint commitment, see Gilbert (2014, 51). Shelby’s analysis appeals to joint commitment (2005, 246), though he may not intend to invoke Gilbert’s notion. On joint commitment and epistemic virtues and vices at the group level, see de Ridder (2022), de Rooij and de Bruin (2022), Fricker (2010), Fricker (2020), Holroyd (2020).
- 14 I argued above that if sharing values involves no more than *de facto* agreement in evaluative beliefs, then condition (1) doesn’t entail condition (2). I suspect that is also true of sharing values on the middle road, whereas if we take the more rigorous road, then condition (1) may entail condition (2). Still, the overall point is that whichever road we take, solidarity will require conditions (1) and (2).
- 15 Relatedly, see Zagzebski’s (1996, 132–134) distinction between motivational dispositions to act in a particular way and behavioral dispositions to act in that particular way.
- 16 There are two ways in which one motivational disposition might be stronger than another: the disposition might be stronger by being more consistent, or the motivation itself might be stronger even though the dispositions are equally consistent.
- 17 See Bommarito (2016, 450).
- 18 On joint and collective action, see Bratman (2014); Gilbert (2014); List and Petit (2011).
- 19 See Fantl (2018, 140 fn20).
- 20 I have argued that if we take the middle road, solidarity requires conditions (4a) and (4b). It will also require these conditions if we take the more rigorous road.
- 21 They will fall short even if they satisfy conditions (1)–(4) above.
- 22 Scholz (2008, 82) argues that this feeling of belonging is not necessary for political solidarity.
- 23 Nor were the members of LGSM card-carrying members of NUM, though they were members of the larger group composed of LGSM and NUM.
- 24 Many virtue theories distinguish between possessing a virtue, on the one hand, and performing the same actions that a virtuous person would perform, on the other. Here, I am applying this to traits (which need not be virtues).
- 25 Compare Medina’s account of solidarity:

a community of subjects who are prepared to think and believe together as they act upon their beliefs through collaborations, and who are ready to be responsive and accountable to each other as they try to share their experiential vantage points and to coordinate their actions.

(2013, 274)

Medina argues that the liberatory epistemic virtue of radical solidarity involves a shared commitment to “acknowledge and respond to ... heterogeneous perspectives” with the shared ulterior aims of correcting biases, and increasing learning, objectivity, and epistemic justice (2013, 277).

- 26 For an application of this specific approach to the trait and virtue of humility, see Whitcomb et al. (2020). This general Aristotelian approach is similar to Tessman (2005, Ch. 6) and MacIntyre (1984), which provide neo-Aristotelian analyses of two virtues similar to solidarity—loyalty and patriotism, respectively.
- 27 Cf. Adams' distinction between motivational and structural virtues (2006, 33–34).
- 28 Though I will be focusing on condition (4), and to a lesser extent (3), presumably group members can exceed with respect to each of the dispositions in (1)–(5).
- 29 See also McKinnon (2017, 170) on the first-person authority of members of disadvantaged groups with respect to discrimination.
- 30 Additionally, the epistemic value of facilitating the virtue of (e.g.) epistemic autonomy may sometimes trump the epistemic value of knowledge.
- 31 On vice epistemology, see, for example, Battaly (2014), Cassam (2019), Kidd et al. (2020).
- 32 Some members of black liberation movements have claimed that prioritizing the sub-goals of black women would minimize or even worsen the oppression of black men.
- 33 hooks (2014) is one place to start.
- 34 But, he restricts these warnings to black political solidarities that require a shared conception of 'thick' black identity.
- 35 See Fricker (2007).
- 36 Cf. Clark (2014) on excesses and deficiencies of solidarity.

References

- Adams, Robert M. 2006. *A Theory of Virtue*. Oxford: Clarendon Press.
- Aristotle. 1984. "Nicomachean Ethics." In J. Barnes (ed.) *The Complete Works of Aristotle*. Princeton, NJ: Princeton University Press.
- Battaly, Heather. 2014. "Varieties of Epistemic Vice." In J. Matheson and R. Vitz (eds.) *The Ethics of Belief*. Oxford: Oxford University Press, 51–76.
- Blum, Laurence. 2007. "Three Kinds of Race-Related Solidarity." *Journal of Social Philosophy* 38(1): 53–72.
- Bommarito, Nicolas. 2016. "Private Solidarity." *Ethical Theory and Moral Practice* 19: 445–455.
- Bratman, Michael. 2014. *Shared Agency*. Oxford: Oxford University Press.
- Broncano-Berrocal, Fernando and J. Adam Carter. 2021. *The Philosophy of Group Polarization*. New York: Routledge.
- Byerly, Ryan and Meghan Byerly. 2016. "Collective Virtue." *Journal of Value Inquiry* 50: 33–50.
- Cassam, Quassim. 2019. *Vices of the Mind*. Oxford: Oxford University Press.
- Cherry, Myisha. 2020. "Solidarity Care." *Public Philosophy Journal* 3(1): 1–12.
- Cherry, Myisha. 2017. "State Racism, State Violence, and Vulnerable Solidarity." In N. Zack (ed.) *The Oxford Handbook of Philosophy and Race*. Oxford: Oxford University Press.
- Clark, Meghan J. 2014. "Anatomy of a Social Virtue: Solidarity and Corresponding Vices." *Political Theology* 15(1): 26–39.
- Coplan, Amy. 2011. "Understanding Empathy: Its Features and Effects." In A. Coplan and P. Goldie (eds.) *Empathy: Philosophical and Psychological Perspectives*. Oxford: Oxford University Press, 3–18.

- Coplan, Amy and Peter Goldie (eds.) 2011. *Empathy: Philosophical and Psychological Perspectives*. Oxford: Oxford University Press.
- Cureton, Adam. 2012. "Solidarity and Social Moral Rules." *Ethical Theory and Moral Practice* 15: 691–706.
- de Ridder, Jeroen. 2022. "Three Models for 'Collective Virtues.'" In M. Alfano, C. Klein, and J. de Ridder (eds.) *Social Virtue Epistemology*. New York: Routledge, 367–385.
- de Rooij, Barend and Boudewijn de Bruin. 2022. "Real Life Collective Epistemic Virtue and Vice." In M. Alfano, C. Klein, and J. de Ridder (eds.) *Social Virtue Epistemology*. New York: Routledge, 369–414.
- Fantl, Jeremy. 2018. *The Limitations of the Open Mind*. Oxford: Oxford University Press.
- Fricker, Miranda. 2020. "Institutional Epistemic Vices: The Case of Inferential Inertia." In I.J. Kidd, H. Battaly, and Q. Cassam (eds.) *Vice Epistemology*. New York: Routledge, 89–107.
- Fricker, Miranda. 2010. "Can There Be Institutional Virtues?" In T. S. Gender and J. Hawthorne (eds.) *Oxford Studies in Epistemology*. Oxford: Oxford University Press, 235–252.
- Fricker, Miranda. 2007. *Epistemic Injustice*. Oxford: Oxford University Press.
- Gilbert, Margaret. 2014. *Joint Commitment*. Oxford: Oxford University Press.
- Holroyd, Jules. 2020. "Implicit Bias and Epistemic Vice." In I.J. Kidd, H. Battaly, and Q. Cassam (eds.) *Vice Epistemology*. New York: Routledge, 126–147.
- hooks, bell. 2014. *ain't i a woman: black women and feminism* (2nd ed.) New York: Routledge.
- Kidd, Ian James. 2020. "Epistemic Corruption and Social Oppression." In I.J. Kidd, H. Battaly, and Q. Cassam (eds.) *Vice Epistemology*. New York: Routledge, 69–85.
- Kidd, Ian James, Heather Battaly, and Quassim Cassam (eds.) 2020. *Vice Epistemology*. New York: Routledge.
- Kolers, Avery. 2016. *A Moral Theory of Solidarity*. Oxford: Oxford University Press.
- Kolers, Avery. 2012. "Dynamics of Solidarity." *The Journal of Political Philosophy* 20(4): 365–383.
- List, Christian and Philip Pettit. 2011. *Group Agency*. Oxford: Oxford University Press.
- MacIntyre, Alasdair. 1984. *Is Patriotism a Virtue?* Lawrence: University of Kansas Press.
- McKinnon, Rachel. 2017. "Allies Behaving Badly: Gaslighting as Epistemic Injustice." In I.J. Kidd, J. Medina, and G. Pohlhaus (eds.) *The Routledge Handbook of Epistemic Injustice*. New York: Routledge, 167–174.
- Medina, José. 2013. *The Epistemology of Resistance*. New York: Oxford University Press.
- Scholz, Sally J. 2008. *Political Solidarity*. University Park, PA: The Pennsylvania State University Press.
- Shelby, Tommie. 2005. *We Who Are Dark: The Philosophical Foundations of Black Solidarity*. Cambridge, MA: Harvard University Press.
- Shelby, Tommie. 2002. "Foundations of Black Solidarity: Collective Identity or Common Oppression?" *Ethics* 112: 231–266.

- Swanton, Christine. 2003. *Virtue Ethics*. Oxford: Oxford University Press.
- Táíwò, Olúfẹ̀mi. 2020. "Identity Politics and Elite Capture." *Boston Review: A Political and Literary Forum*. <http://bostonreview.net/race/olufemi-o-taiwo-identity-politics-and-elite-capture>. Accessed: May 31, 2021.
- Tessman, Lisa. 2005. *Burdened Virtues*. Oxford: Oxford University Press.
- Thomas, Laurence. 1993. "Moral Flourishing in an Unjust World." *Journal of Moral Education* 22(2): 83–96.
- Whitcomb, Dennis, Heather Battaly, Jason Baehr, and Daniel Howard-Snyder. 2020. "The Puzzle of Humility and Disparity." In M. Alfano, M. Lynch, and A. Tanesini (eds.) *The Routledge Handbook of the Philosophy of Humility*. New York: Routledge.
- Zagzebski, Linda T. 1996. *Virtues of the Mind*. Cambridge: Cambridge University Press.

T&F Proofs – Not for Distribution

10b Commentary from T. Ryan Byerly

Comments on Battaly

Heather Battaly's chapter provides a detailed account of the collective trait of solidarity. It explains the conditions under which solidarity is a virtue; it highlights how intellectual virtues and vices may be related to maintaining or undermining solidarity; and it discusses the question of whether the example of solidarity provides a challenge to summativism about collective virtues. I'll offer critical comments on two issues arising in the chapter.

The first comment concerns Battaly's arguments for thinking that the trait of solidarity is not always a virtue. Battaly provides two kinds of examples that are supposed to show that solidarity is not always a virtue. First, there are bad groups with bad motives that nonetheless have the trait of solidarity; for them, the trait is not a virtue because of their bad motives. Second, some groups take solidarity to the extreme, with some group members becoming excessively deferential or trusting toward others; in these cases, we have excessive solidarity, which is a vice rather than a virtue.

While I think there is something to these arguments, I also think they could be resisted. First, in the case of bad groups with bad motives, we might still think that these groups can possess some virtues while being criticizable or vicious in other respects. We might think, for instance, that a gang of thieves is reprehensible for their disrespect of others' property or for their greed, but nonetheless admirable for the kind of togetherness and teamwork exemplified in their solidarity with one another. We might think their solidarity in itself is good and even virtuous, while these other features are enough to explain what is problematic about the group (cf. Roberts 1984).

There are some further details that can be supplied to make this case more plausible. First, as briefly alluded to, we might admire the solidarity of bad groups, despite their badness. If we do, then an exemplarist approach to virtue theory may provide some evidence that their solidarity is virtuous (Zagzebski 2017). Second, we might think that in any case in which a group genuinely possesses solidarity the group members will

share some good values such as sharing life with each other, looking out for each other, and working together with one another. If the members of the group value these things *only* instrumentally in order to promote bads such as stealing or feeding their greed, we might think that the group doesn't in fact possess solidarity, as the conditions of their cooperation and mutual concern are too flimsy. If this is indeed correct, then it is more plausible that if bad groups possess genuine solidarity, it will be a virtue for them and not a vice.

Regarding cases of "excessive" solidarity, what I want to suggest is that Battaly may be treating vices of excess differently than many virtue theorists would be inclined to. It is a commonplace of virtue theory that in at least many cases, virtues hit a mean between vices of excess and vices of deficiency. Courage is a mean between cowardice and brashness; generosity is a mean between stinginess and prodigality. Now, virtue theorists will often resist the idea that in these cases the vice of excess is an excessive form of the same trait that is the virtue. For instance, they may take pains to emphasize that brashness is *not* courage, and that prodigality is *not* generosity. Yet, Battaly seems to want to do the opposite with solidarity. She wants to claim that excessive solidarity is solidarity in excess, rather than *not* being solidarity.

The distinction here, however inconsequential it might appear at first glance, does seem to reflect two fundamentally different ways of thinking about the individuation of traits. For Battaly, traits are thin or hollow in a certain way. The same trait can be possessed out of very different motives, and accompanied more broadly by a quite different underlying psychology. Yet, according to another approach to trait individuation, what makes traits the traits they are is precisely their characteristic psychology, including their characteristic motivations (cf. Baehr 2011). While there is much more debate to be had about this topic, I'll briefly note that the increasingly popular whole trait theory and other similar models of personality and character are supported by a growing body of empirical evidence that would seem to favor an approach to trait individuation that is opposed to Battaly's in that it treats motivations and other characteristic psychological features as definitive of traits (see Fleeson and Jayawickreme 2015).

My second comment concerns Battaly's discussion of whether the case of virtuous solidarity provides a counterexample to summativism about group character traits. Summativist accounts of group character traits are standardly defined as claiming that for a group to possess a character trait C is for a sufficient number of the group's members to possess C (Lahroodi 2019). Summativism has been criticized mainly on the basis of cases in which there is divergence between group character and group member character because group members behave in markedly different ways in the group context than when acting on their own behalf as private individuals. For instance, cases of this kind can occur where groups

possess a character trait because members act in accordance with that trait in the group context, while they wouldn't do so outside of the group context.

In previous work (2016), Meghan Byerly and I had suggested an additional way of challenging summativism. We suggested that there may be distinctively collective virtues—virtues that collectives can possess but individuals cannot. If there are, then these will be cases that falsify summativism—cases where a group's possession of a trait T does not consist in enough of its members possessing T. We suggested that solidarity may be an example of a distinctively collective character trait, and we proposed that there is unique work to do in collective virtue theory when it comes to such virtues because accounts of them cannot be straightforwardly transferred from the case of individual virtue analogs.

In my view, Battaly's chapter beautifully illustrates this point. Battaly carefully, systematically builds an account of group solidarity, and does so not by transferring some account of individual solidarity to the group case, but by attending carefully to the group case. This is precisely the sort of work Meghan and I were envisioning.

But Battaly worries that claiming that cases like this pose a threat to summativism may be some kind of "cheating." This is because her account of solidarity "derive[s] a group-level trait and virtue from the traits and virtues of its individual members." I worry, however, that this description of the account is misleading. The account doesn't derive a group-level virtue from virtues of individual members; instead, it derives an account of a group virtue from the characteristic patterns of behavior of group members *acting as group members in the group context*. Indeed, the idea that solidarity is a matter of group members' typical behaviors and attitudes in the group context with respect to other group members and pertaining to shared group goals is front and center in Battaly's account. Reference to the group is ineliminable from her analysis. This is already sufficient to make the case a clear counterexample to summativism. The account may seem to be in the "spirit" of summativism in the sense that the ultimate explanation of the group's solidarity is to be found in patterns of group-oriented behavior of the members. But as far as I am aware anti-summativists were never committed to denying that.

References

- Baehr, Jason. 2011. *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford: Oxford University Press.
- Byerly, T. Ryan and Meghan Byerly. 2016. "Collective Virtue." *Journal of Value Inquiry* 50, 1: 33–50.
- Fleeson, William and Eranda Jayawickreme. 2015. "Whole Trait Theory." *Journal of Research in Personality* 56: 82–92.

- Lahroodi, Reza. 2019. "Virtue Epistemology and Collective Epistemology." In *The Routledge Handbook of Virtue Theory*, ed. Heather Battaly, 407–419. New York: Routledge Press.
- Roberts, Robert. 1984. "Will Power and the Virtues." *Philosophical Review* 93, 2: 227–247.
- Zagzebski, Linda. 2017. *Exemplarist Moral Theory*. Oxford: Oxford University Press.

T&F Proofs – Not for Distribution

10c Commentary from Duncan Pritchard

On Solidarity: Collectivity, Trust, and Deference

For the most part, I agree with Battaly's excellent treatment of solidarity in this chapter, so my critical comments are relatively minor. Given space constraints, I want to focus on three points that I hope will be helpful. The first concerns the idea that Battaly opens the chapter with, that solidarity is a distinctively collective virtue. The second is the claim that solidarity demands trust. The third is the thesis that solidarity essentially involves deference. I will take these points in turn.

What does it mean to say that a virtue like solidarity is a distinctively collective virtue? Attributing this idea to Ryan Byerly and Meghan Byerly, Battaly argues that "a *distinctively* collective virtue is a virtue of a collective (or group) for which there is no individual analog; that is, there is no corresponding virtue *V* of individuals, from which the collective version of *V* could be derived." Solidarity is meant to be a "paradigm" example of a distinctively intellectual virtue in just this sense. While Battaly raises some issues for this way of thinking about solidarity at the end of her piece, she nonetheless broadly endorses it. In contrast, I do not find it convincing.

In particular, while it is obviously true that the virtue of solidarity can only be manifest in a social setting, rather than individually (one cannot manifest solidarity with oneself), I don't think it follows from this that the virtue in question must thereby be a collective virtue, and hence that there is no (theoretically independent) individual virtue in play. The obvious sense in which the virtue of solidarity is social is that solidarity is a virtue that one manifests towards others rather than oneself. But it doesn't follow from that point alone that solidarity is not an individual virtue. Indeed, aren't most virtues essentially other-directed in their manifestation? Consider, for example, the virtue of kindness. To manifest this virtue surely involves appropriately manifesting kindness to others. But we would not conclude on this basis that there is no individual virtue of kindness, much less that this virtue is in fact a collective virtue that couldn't be understood in terms of the individual virtue. So

while I agree that solidarity is an inherently social virtue, I am not convinced that we should regard it as a distinctively collective virtue in the sense described.

Battaly argues, following Tommie Shelby, that solidarity demands trust. In particular, she maintains that it demands that one “be disposed to trust the testimony of other group members with respect to their goals and actions.” Battaly takes this point about trust as following from the fact that solidarity requires acting in accordance with shared goals, as this “will entail knowing in broad outline that one is working with others who have the same goals, knowing roughly what kinds of actions they are performing in pursuit of those goals, and coordinating one’s actions with theirs.” Crucially, however, she further claims that we cannot gain this knowledge unless “we trust the testimony of our fellow members; knowing these things requires trusting their testimony with respect to their goals and their actions.”

I’m not convinced that solidarity does require trust. Can’t there be solidarity amongst a criminal family for example? If so, then it is hard to see why solidarity requires trust. Presumably Battaly would claim that in this case what’s on display is not solidarity but rather mere loyalty, which she claims doesn’t demand trust. But since the only reason we are given for thinking that loyalty and solidarity come apart is that only the latter demands trust, a natural response would be to maintain that in fact *neither* of them requires trust. Absent any additional supporting argument, why aren’t members of the criminal gang displaying solidarity rather than just loyalty?

In any case, I’m not convinced by the reason Battaly gives for claiming that solidarity demands trust. It simply isn’t true that the only way to gain knowledge of shared goals is by trusting the word of others, as one can usually gain independent grounds for the target beliefs, and hence one needn’t simply rely on the other person’s say-so. In any case, there seem to be genuine instances of solidarity where there isn’t trust and where, therefore, knowledge of shared goals is not based on trust. Consider, for example, the solidarity of the members of a group of recovering addicts (at an AA group, for instance). Recovering addicts know full well the dangers of trusting the word of fellow recovering addicts, especially concerning anything related to the issue of their addiction, and so would be naturally circumspect about each other’s testimony in this regard. But that needn’t prevent them from coming to know enough about their shared goals in order to act in accordance with them. Solidarity thus doesn’t seem to demand trust, nor is trust in any case required in order to have the kind of social knowledge required for the manifestation of solidarity.

This brings us to a final point, which concerns solidarity and deference. Battaly insists that solidarity demands deference, where this is itself a kind of trusting of other group members. She takes this point to

follow from the requirement that solidarity requires group action. She writes that in order “to be disposed to coordinate our actions with the group’s, when our judgments are initially in conflict with the group’s, we must be disposed to at least sometimes trust, and defer to, the group’s judgment.” That is, if we didn’t defer in such circumstances, then we would be acting contrary to the group, and hence no longer coordinated with it. Hence, deference is required.

I don’t find this line of argument very convincing. To begin with, notice that it depends on an implausibly robust conception of the kind of coordination of one’s actions with the group that solidarity demands. Battaly earlier outlines this requirement in terms of the convincing claim that one needs to be disposed to act in accordance with the shared goals and values of the group, but given that this is a general disposition it doesn’t preclude one being willing to sometimes act contrary to the goals of the other members of the group. General dispositions are not exceptionless, after all. Moreover, there seems to be a salient rationale for an exception in play here, which is when one is clearly in a state of elevated knowledge and expertise relative to the group. In that case, why should one be disposed to defer to the group? To take the example of domestic violence that Battaly uses in this regard, suppose that one has suffered for years as a victim of domestic violence (and so one has first-hand experience of the phenomenon) and then one goes on to become a recognized authority on domestic violence (and so one has an exceptional level of relevant theoretical knowledge too). It would now seem entirely appropriate to not defer to the group, and indeed to be willing to not align with the group’s actions where one judges that these actions are based on poor judgment. Lack of deference here seems entirely compatible with solidarity.

10d Heather Battaly's Response to Commentaries

Replies to Byerly and Pritchard

HEATHER BATTALY

I am delighted and grateful to be thinking through these issues with Byerly and Pritchard and to be practicing social epistemology with them! Together, their commentaries raise four important sets of questions: (1) Is solidarity a distinctively collective virtue? (2) Can't solidarity still be a virtue of "bad actors" with bad motives? (3) What are the connections between traits, virtues, and vices of excess? (4) Does solidarity require trust? (5) Does it require deference? Below, I'll briefly sketch some initial replies, and flag some open questions that merit further exploration.

The first set of questions focuses on antismmativism and what it means for a virtue to be distinctively collective. Byerly and Pritchard both engage the issue of whether solidarity is distinctively collective: Pritchard argues that it isn't; conveniently, Byerly argues that it is! Here, I'll focus on Pritchard's worry. He suggests that solidarity is an other-regarding virtue like kindness, which can clearly be possessed by individuals (as well as groups) and thus isn't distinctively collective. Whereas, I have argued that possessing the virtue of solidarity requires possessing the trait of solidarity, which irreducibly involves relations among individuals and can only be possessed by a group that has individuals as members. Granted, we do sometimes *say* of an individual that they are *in* solidarity *with* a group (or another individual), and my view may need an error theory for this. Briefly, I suspect we use this language as shorthand for saying that the individual has some of the same values as the group and/or coordinates their actions with the group's. Notice, however, that we don't tend to say that an individual *has* solidarity (full stop). An individual may well believe and act *in* solidarity *with* a group, but lone individuals don't possess the trait of solidarity, and thus don't possess the virtue. In contrast, lone individuals can and do possess the trait and virtue of kindness, where this consists in something like a disposition to care about and aid others. Byerly and Pritchard's exchange demonstrates that we need clear distinctions between other-regarding virtues that can be possessed by individuals and distinctively collective

virtues that cannot. If solidarity is a trait and virtue that can only be possessed by groups that have individuals as members, we also need analyses of group membership.

Second, Byerly suggests that solidarity can still be a virtue of “bad actors” with bad motives. One of his points references Roberts (1984), which distinguishes between virtues of willpower and motivational virtues. If I am understanding Byerly here, he is suggesting that the virtue of solidarity, like virtues of willpower, might not require good motives. His worry raises tricky theoretical issues both about the structure of virtues and about the connections between ulterior and proximate motives. In partial reply, I am a pluralist about virtue insofar as I think there is more than one kind of virtue and more than one way for a trait to be valuable. Roughly, motives-virtues derive their value at least partly from intrinsically good motives, whereas effects-virtues derive their value from the good effects they produce (Battaly 2015). The same trait can be a motives-virtue and an effects-virtue. Above, I argued that solidarity is a motives-virtue. I think it can also be an effects-virtue. That said, I have difficulty seeing how solidarity could be a virtue, when it is driven by bad motives and produces bad effects. In other words, why wouldn't solidarity in “bad actors” instead be a vice? What exactly is supposed to be valuable about the solidarity of “bad actors”?

Third, Byerly raises concerns about my “normative contextualism” (Kidd 2020)—my assumption that there is a normatively neutral trait of solidarity, which can sometimes be a virtue and sometimes be a vice. Byerly follows many virtue theorists in assuming that our concepts of solidarity, courage, etc. are normatively thick. If we follow that common assumption, then brashness, which is a vice of excess, won't be an excess of courage, it will be something other than courage. Likewise, group-think, which is a vice of excess, won't be an excess of solidarity, it will be something other than solidarity. In reply, normative contextualism can explain what is *excessive* in vices of excess like brashness and group-think—the normatively neutral traits of courage and solidarity are excessive—and thus can explain why these vices count as excesses of courage and solidarity, respectively. These vices are excesses of the respective traits, not the respective virtues. Whether this gives normative contextualism an explanatory advantage over normatively thick accounts remains to be seen and is well worth exploring.

Fourth, Pritchard argues that solidarity doesn't require trust, since we can gain knowledge of whether or not group members share (and act in accordance with) the group's goals via independent means and without relying on their testimony. Further, he suggests that there can be solidarity among, for example, members of a group of recovering addicts even when they don't trust one another's testimony. These points highlight a helpful question: does solidarity require trusting the testimony of a specific individual member of a group with respect to their *own* actions

and goals (4a)? Suppose we were to ultimately decide that it doesn't. Even so, given that (i) large groups (members of a global NGO) can have solidarity, and that (ii) solidarity requires knowing about the goals and actions of fellow group members in rough outline, and, further, that (iii) we won't be in a position to gain such knowledge via independent means in large groups, solidarity would still require trusting the testimony of someone or other, at least pragmatically. Could that include trusting the testimony of non-members, and is that a problem for the account? These questions show that we need further exploration of whether and why trust is required for solidarity.

Pritchard likewise seems to suggest that solidarity needn't require even occasional deference to the group's judgment, when one is the only expert in the group. Briefly, in reply, knowledge isn't the only thing of epistemic value and isn't the only factor relevant to appropriate deference. Growth in intellectual virtues, such as intellectual autonomy, is also epistemically valuable and relevant to appropriate deference. I suspect the virtue of solidarity will sometimes require experts to go along with the actions recommended by the group—even though they know the group is unreliable in judging which actions are effective—in the interests of facilitating the intellectual autonomy of group members. Clearly, the virtue of solidarity won't always require this.

While I can't pretend to have done their points justice, I am grateful to Byerly and Pritchard for their incisive responses.

References

- Battaly, Heather. 2015. *Virtue*. Cambridge: Polity Press.
- Kidd, Ian James. 2020. "Epistemic Corruption and Social Oppression." In I.J. Kidd, H. Battaly, and Q. Cassam (eds.) *Vice Epistemology*. New York: Routledge, 69–85.
- Roberts, Robert C. 1984. "Will Power and the Virtues." *Philosophical Review* 93(2): 227–247.

11 Collective ('Telic') Virtue Epistemology

J. Adam Carter

1 Introduction

A familiar theory of individual propositional knowledge holds that propositional knowledge is type identical with *apt* belief. A belief is apt if and only if it is successful (i.e., accurate) because competent.¹

When suitably fleshed out, this view has a lot of explanatory power in individual epistemology. It can, among other things, help us navigate the Gettier problematic,² the Pyrrhonian problematic,³ radical sceptical challenges,⁴ the value problem,⁵ and more recently, epistemological problems related to the suspension of judgement.⁶

What's much less clear—and almost entirely unexplored⁷—is whether the 'knowledge = apt belief' (K = AB) template view is applicable only in individual epistemology, or whether some version of it can be made to work in *collective epistemology* as well, as a thesis about *group knowledge*.

One way in which things get messy here is as follows. The (K = AB) view is able to get all of the pleasing results above only when it is understood against a particular background view of epistemic normativity as *telic normativity*—viz., the normativity of attempts *as* attempts.⁸ According to this background view, X-attempts can be evaluated along three dimensions: for success (did the attempt succeed in attaining X), for adroitness (did the attempt manifest a competence to attain X reliably enough) and for aptness (was the attempt successful because adroit?). *Believing* is a kind of attempt, one that aims constitutively at getting it right (i.e., at truth).⁹ Accordingly, the (K = AB) thesis goes hand in hand with the idea that only those beliefs that enjoy a certain kind of normative assessment *qua* attempt—the status of being not merely successful and adroit, but also *apt*—qualify as known.

Already, though, with the above qualifications in play, there is a serious barrier to getting any kind of 'knowledge = apt belief' view off the ground in collective epistemology, as view of collective knowledge. First, as recent literature in the epistemology of groups suggests,¹⁰ even if groups can possess *knowledge*, it is much less clear that groups can have *beliefs*, and thus, that they can make the *kind of attempt* that, on

the (K = AB) view, aspires to knowledge. Secondly, even those collective epistemologists who *do* countenance collective beliefs often (though not always) take such collective beliefs to be a function of *joint acceptance* of a proposition—viz., whereby a group jointly accepts a proposition, $\langle p \rangle$, if and only if its members commit to acting as if $\langle p \rangle$ is true in their capacity as group members.

But here is the fly in the ointment: individual beliefs and collective joint acceptances seem, *prima facie*, like very different *kinds of attempts*. The latter, after all, involves intentional action in a way that is broadly analogous to how individual-level acceptances are intentional. But, individual-level acceptances are not governed by a true aim. Accepting a proposition, in individual epistemology, is not unsuccessful if untrue; put another way, acceptance is not a constitutive attempt at truth. The problem now sharpens: if individual belief and collective joint acceptance are not normatively constrained by the same aim, then (K = AB) will not be extensionally adequate at the collective level as a theory of group knowledge, even if it delivers the goods at the individual level.

2 (K = AB) and Substantive Symmetry

In light of the above, let's consider now the salient option space that the (K = AB) proponent at the individual level has for doing collective epistemology.¹¹

<i>View</i>	<i>Individual View/Collective View</i>
<i>Collective knowledge scepticism</i>	K = AB (non-sceptical) / K=AB (sceptical)
<i>Non-sceptical hybrid traditionalism</i>	K = AB (non-sceptical) / not K=AB and not K-First (non-sceptical)
<i>Non-sceptical hybrid knowledge-first</i>	K = AB (non-sceptical) / not K=AB and K-First (non-sceptical)
<i>Non-sceptical symmetric view</i>	K=AB (non-sceptical) / K=AB (non-sceptical)

Each of these four options comes with some substantial *prima facie* costs, though some to a greater extent than others. Option 1, *collective knowledge scepticism*, submits that if there is collective knowledge, (K = AB) would be a correct theory of that knowledge; however, since 'AB' is on this view held to be not realisable at the collective level, there simply is no collective knowledge. *Collective knowledge scepticism* effectively throws in the towel in collective epistemology, relegating knowledge (and thus epistemology) entirely to the individual arena. The cost here is significant, given, for instance, the prevailing view in collective epistemology that there are at least some *bona fide* cases of distributed knowledge, viz., cases where a group knows something that is not reducible to any proposition known by any of the individual members of the group.¹²

Option 2, the *non-sceptical hybrid traditional view*, cobbles together the (K = AB) view at the individual level with some other reductive—viz., *non*-‘knowledge-first’—theory of knowledge at the collective level (e.g., perhaps collective knowledge = justified, true, collective acceptance), and maintains that the conditions specified by this other theory are (ordinarily enough) collectively realised. This view seems to do better than *collective knowledge scepticism* in one sense but much worse in another. It does better in the sense that it does not throw in the towel at the collective level. But the cost of the non-sceptical result here is foregoing an important theoretical desiderata, *substantive symmetry*, which we can define as follows:

Substantive symmetry: A theory of individual and collective knowledge, T, is substantively symmetrical if and only if T posits a description of knowledge conditions at the collective level that matches T’s description of knowledge conditions at the individual level; otherwise, T is substantively asymmetrical (viz., ‘hybrid’).

Notice that Option 2, the *non-sceptical hybrid traditionalist view* is clearly substantively asymmetrical. This substantive asymmetry across the individual/collective divide invites a metatheoretical objection: when a view is substantively *asymmetrical*, this substantive asymmetry is (defeasible) evidence that the view is identifying something other than knowledge at least one of the two levels. And this general metatheoretical objection (which requires further explanation on the part of the view to address it) gains additional traction once attention is drawn to the *prima facie* putative differences between individual and collective believing, and in particular, to the intentional character of the latter and the non-intentional character of the former.

Option 3, the *non-sceptical hybrid knowledge-first view*, avoids a sceptical implication in collective epistemology. That’s good. And it also sidesteps entirely complications that arise for views—like Option 2—that attempt to vindicate collective knowledge by showing how it can be dismantled into a collective belief (or acceptance) condition plus other epistemic conditions that are satisfied if and only if the collective has knowledge. In recent co-authored work,¹³ I’ve argued that—as a view of collective knowledge—a K-first view¹⁴ has much to recommend it over ‘justified true collective belief’ and ‘justified true collective acceptance’ accounts of collective knowledge.¹⁵

There remains, though, the lingering issue of substantive symmetry. Going ‘K-first’ at the collective level achieves substantive symmetry in individual and collective epistemology *only if* it is paired with a K-first view at individual level. My co-authors with whom I’ve recently written on behalf of K-first collective epistemology—Christoph Kelp and Mona Simion—are happy to embrace knowledge-first epistemology at

the individual level.¹⁶ What this means is that by going K-first at the collective level, they maintain substantive symmetry, whereas I would maintain this only by then jettisoning (K = AB) at the individual level. As the reader will have gleaned from §1, I think it is right to be impressed with what (K = AB) can do at the individual level, and more so than with what K-first can do.

For those epistemologists who like me favour (K = AB) at the individual level, then, the prospects don't look initially very good to get 'everything we want' at the collective level, including substantive symmetry. *Unless* of course, there is some way to make Option 4, the *non-sceptical symmetric view* work, *despite* all of the worries noted in §1 that seem to stand in the way of defending (K = AB) at the collective level.

In the remainder of this chapter, I am going to argue that—despite things looking very bleak initially—Option 4 really *is* defensible. (K = AB) can be made to work not only as a theory of individual knowledge but *also* as a theory of collective knowledge. In order to see how the view works, it will be important to distinguish, following Ernest Sosa (2015, 2020), two importantly different kinds of beliefs: mere *alethic affirmations* and *judgemental beliefs*. While there is (perhaps) no collective analogue to individual alethic affirmations, there is a collective analogue to individual judgemental beliefs. And—here is the second part of the argument—the collective analogue of individual judgemental beliefs is performatively analogous with individual level judgemental beliefs, and in this respect, it is analogously knowledge-apt. By drawing a parallel between individual judgement and collective judgement, we can see how (K = AB) is not merely a serious option for individual epistemology but for collective epistemology as well.

3 Grades of Knowledge

Following Sosa, the (K = AB) slogan can be unpacked in individual epistemology in different ways—corresponding with different *grades of knowledge*—depending on the kind of *attempt* at getting it right that the relevant belief is.

Suppose, for example, that you are taking an eye exam, and you begin losing confidence as you get closer to the bottom row. But you read out the bottom row anyway. It turns out your lack of confidence on those bottom-row letters wasn't warranted, as you were actually perfectly reliable at the bottom row, despite the shaky confidence.¹⁷

When you affirm a given letter, $\langle p \rangle$, on the bottom row, it is an attempt to get it right (on whether $\langle p \rangle$) by affirming that $\langle p \rangle$ —viz., it is, in Sosa's terminology, a kind of *alethic affirmation*. In the above case, your alethic affirmation that $\langle p \rangle$ is *adroit* (more so than you recognise); and even more, your alethic affirmation that $\langle p \rangle$ is *apt*. As such, it constitutes a kind of subcredal *animal knowledge*, where animal knowledge is type-identical with *apt alethic affirmation*.¹⁸

So: ‘animal knowledge = apt alethic affirmation’ represents one kind of substantive gloss of the (K = AB) template. But, importantly, not *all* beliefs are *mere* alethic affirmations, viz., attempts to simply get it right by affirming. Some beliefs—*judgemental beliefs*—are attempts to get it right *aptly* by *alethically affirming*. In a bit more detail for now (see §4 for further elaboration): in judging something to be so, a thinker aims intentionally to get it *aptly* right (by alethically affirming that $\langle p \rangle$) on a given question. *Apt judgement*—that is, when one’s attempt to get it right aptly by alethically affirming is *itself* apt—is knowledge. But it’s not merely the kind of knowledge you get when a mere alethic affirmation is apt. Apt judgement is—on the telic virtue epistemologist’s framework—knowledge *full well* (alternatively: judgemental knowledge).

In sum, then, the (K = AB) template account of knowledge can be glossed—within telic virtue epistemology—in the following two ways, which correspond with two different kinds of ‘attempts’ to get it right through affirmation.

What kind of attempt?	What does it constitutively aim at?	What is it when apt?
alethic affirmation	to get it right (whether $\langle p \rangle$) by affirming that $\langle p \rangle$	animal knowledge
Judgement	to get it right (whether $\langle p \rangle$) aptly by alethically affirming that $\langle p \rangle$	judgemental knowledge (i.e., knowledge full well)

As the reader might have anticipated, judgemental knowledge is of particular interest for the telic virtue epistemologist who wants to embrace a (K = AB) view at both the individual and collective level, and thus, to retain substantive symmetry at the individual and collective levels. And this is because judgement, *qua* attempt, involves an intention—where the intention here is to get it right *aptly* by alethically affirming. Intention, recall, seemed *ex ante* to be a property of collective belief (or collective acceptance) not shared with individual belief. Even more, and unlike mere acceptance, the constitutive aim of judgement is such that, when aptly attained, what results is a kind of knowledge.

The takeaway point, then, seems to be the following: a proponent of (K = AB) at the individual level has a real shot at vindicating a *non-sceptical symmetric view* (§2) despite what looked like initial barriers. The key will be to set aside the collective analogue of mere alethic affirmation and to instead zero in on the collective analogue of individual intentional judgement.

Here is the plan for the remainder of the chapter. §4 will unpack some of the key features (glossed over so far) of judgemental belief at the individual level, and in doing so clarify the sense in which judgemental belief is a species of intentional action on the telic VE framework. §§5–6 then transpose individual judgement and judgemental knowledge (and its key telic normative features) to the collective level, establishing a proof of concept of

how the (K = AB) template can be vindicated as a theory of (judgemental) knowledge at the collective level, and one whose conditions are ordinarily enough met so as to be a non-sceptical theory. §7 then shows that the view has important advantages over a rival version of collective virtue epistemology defended in recent work by Jesper Kallestrup (2020).

4 Judgement

On telic virtue epistemology, a judgement that $\langle p \rangle$ is an intentional constitutive attempt to get it right aptly by alethically affirming that $\langle p \rangle$. In order to bring into view what a collective analogue of an *apt* judgement would be, let's first sharpen some of the components of individual judgement itself, by briefly answering some key 'FAQs':

- (a) In what sense is a judgement an *attempt*?

There are two ways one might attempt to attain an objective, corresponding with a distinction between *instrumental attempts* and *constitutive attempts*.¹⁹ In the former case, one makes an attempt, at an objective, *O*, by implementing means that are both preliminary and viewed as such. Inquirers, for example, might instrumentally attempt to know whether there is a chaffinch in the garden by implementing the preliminary means of finding a good vantage point from which to spot the bird. *Constitutive attempts* are different. In the case of a constitutive attempt at an objective *O*, one implements means aimed at *O*, but not means that are regarded as preliminary and viewed as such. Rather, one—in making a constitutive attempt at attaining *O*—implements means that are aimed at grounding one's success in attaining *O*. (In an athletic case, compare: instrumentally attempting to hit a hole-in-one by practicing hard for months, and constitutively attempting to hit a hole-in-one by swinging the club.)

The idea that a judgement is a constitutive attempt to get it right whether $\langle p \rangle$ aptly by alethically affirming that $\langle p \rangle$ registers this constitutive rather than instrumental character of the attempt.

- (b) So judgements are constitutive attempts. But if the idea is that they are constitutive attempts at *apt* alethic affirmation (and that they are not merely constitutive attempts at *successful* alethic affirmation), then does that mean that judgements are attempts to do more than just 'get it right' (whether $\langle p \rangle$ by affirming' that $\langle p \rangle$?

Yes, that's right. And a helpful way to think about this will be to consider two ways you might constitutively attempt to make a basketball shot. You might on the one hand shoot in the endeavour of *making* it by shooting it. On the other hand, you might shoot in the endeavour of *making it aptly* by shooting it. A reckless shot from too far out—viz.,

beyond your threshold for sufficient reliability—that happens to go in is successful relative to the first kind of attempt. It is not successful relative to the latter kind of attempt, even though it goes in. The basketball coach will advise a player to take only the second kind of shot (unless the clock is running out). The second kind of shot, which aims not just at success, but at aptness, is analogous to a judgement, which aims not just at getting it right, but at getting it right aptly.

- (c) Even if all of the above about judgement is granted, there remains an elephant in the room. Why should we think that the kind of constitutive attempt at apt alethic affirmation that judgement is should be understood as a constitutive *and intentional* attempt at attaining this aim? After all, judgements, like many other beliefs we have, surely aren't voluntary?

To say that judgement is a (constitutive) attempt, with *intention*, to attain a given aim (to wit, the aim of apt alethic affirmation, or animal knowledge) does not imply that how we judge is thereby under the sort of voluntary control whereby we could judge directly through arbitrary choice.²⁰

In order to see why, let's consider what suffices for the kind of intentional action that judgement is. To do this, we can distinguish between a basic action and a 'simple' intentional action. Suppose I intentionally move a finger. Or think of a triangle. These are both deeds I do intentionally, but in each case, there's no other deed I do in the endeavour to do these things, either the physical basic action of moving a finger or the mental basic action of thinking of a triangle.

Affirming that $\langle p \rangle$ is a basic action. When one affirms that $\langle p \rangle$, one doesn't do this partly by doing something else in the endeavour to affirm that $\langle p \rangle$. Basic action can be distinguished from 'simple' intentional action, where an agent aims to perform a deed (at t) at least partly by performing a basic action, B , at t .²¹ Whereas affirming is a basic action, judgement is not; it is a simple intentional action. Here's Sosa (2015):

In a judgment, the agent affirms in the endeavor to (thereby) affirm aptly. If the agent does attain that objective, then, we have the following structure: the agent affirms aptly that p at t partly-by affirming that p at t .

(2015, 166)

The above sense in which judgement is a kind of (simple) intentional action, thus, doesn't imply at all that judging is voluntary in the sense that it could be reversed arbitrarily—viz., the sense in which researchers widely deny that belief, more generally, is voluntary. Moreover, given that not all beliefs are judgemental, not all beliefs are intentional even in the above sense that does not imply voluntary control.

- (d) But if the judgement is intentional even in the sense described, then couldn't it potentially be redirected towards practical ends in a way that is analogous to how individually accepting a proposition can be practically aimed? If the answer is 'yes', then wouldn't 'apt judgement' be a candidate for knowledge if and only if apt acceptance is a candidate for knowledge?

Careful! There's a sense in which judgements are beholden to practical factors: we might let practical values dictate which inquiries we take up in the first place—the kind of normativity here (whereby our doing so is better or worse) is the (broader) normativity of intellectual ethics.²² For example, the lawyer—in the course of considering the case of a potential client—might make various judgements about how the law applies in that particular case, all broadly for the sake of a moral objective: to provide assistance to a vulnerable client. But, in making these judgements, the lawyer is nonetheless making a constitutive attempt at knowing with one's alethic affirmation. Furthermore, the telic assessment of a judgement—the kind of assessment that matters for whether the judgement is knowledge—is not, as Sosa (2020, 36) puts it, 'properly affected by extraneous objectives that the agent may also be pursuing through the same means'.

The above kind of case is very different from constitutively attempting to do something *else* (other than to get it right aptly) by affirming, even when that something else is epistemic. For example, a contestant on a game show who affirms that $\langle p \rangle$ in the endeavour to get it right whether $\langle p \rangle$, by affirming, fully cognisant *that* they are guessing, is not thereby judging. They are merely alethically affirming, making a constitutive attempt to get it *right* (not: aptly right) by affirming. Likewise, one is not judging if one is engaging in wishful thinking—affirming what one merely hopes is the case, or, when affirming through speech, what one merely hopes others believe that one believes is the case.

5 Collective Judgement

With the key contours of judgement, at the individual level, in mind—let's now transpose things to the collective level. As was noted in §1, it is contentious in collective epistemology, and more widely in collective intentionality, whether groups can have genuine *beliefs*, or whether they can merely *accept* propositions.

A quick clarification is in order. There is no dispute whatsoever that groups can have beliefs in an uncontentious *summative* sense. A group summatively believes a proposition, $\langle p \rangle$, if and only if most or all of its members believe that $\langle p \rangle$. For example, the group that is 'Swedes' believes the proposition $\langle \text{Volvos are safe} \rangle$ in a summative sense if and only if all or most individual Swedes believe this.

The more contentious, and more philosophically interesting, notion of group belief is a non-summativist, or alternatively an ‘inflationist’, notion of group belief, according to which it’s false that the group belief is reducible to an aggregate of the individual beliefs of group members.

As Alexander Bird (2019) notes, the rationale that has persuaded those collective epistemologists who have embraced an inflationist view of group belief is that the conditions specified by the summativist account seem in certain cases neither sufficient nor necessary for a group to believe something. Here is an example that challenges sufficiency. Suppose all members of a group (say, a city council) believe $\langle p \rangle$ but, in their capacity as council members behave as though they do not believe $\langle p \rangle$ (and, where each member thinks they are the only one to have this ‘strange’ belief that $\langle p \rangle$.) It seems like, in this case, it would be a mistake to say the city council believes that $\langle p \rangle$ even though all its members do.²³ Likewise, regarding necessity—suppose a group of jurors is evaluating the guilt of an immigrant whose ethnicity is different from their own. The evidence for innocence is overwhelming, and the jurors—each of whom is privately racist and believes the defendant is guilty—recognise this, and publicly affirm in line with the evidence (rather than in line with their racist beliefs) by voting for his innocence. Here, it looks like the *jury* believes the client is innocent, even though its individual members do not. Likewise, cases of distributed cognition speak against necessity: a group might come to endorse a viewpoint, as a group, by way of a division of cognitive labour dispersed across group members, who have different cognitive roles and then ‘feed’ different bits of information to a centralised database that collates the information.

As was noted in §1, the matter of whether we ought, in doing epistemology, to regard these kinds of cases as ones that feature (non-summativist) group *belief* has been deeply divisive. The divisiveness has centred around two claimed disanalogies between belief on the one hand, and so-called inflationary group ‘belief’ on the other. These disanalogies concern (i) automaticity; and (ii) involuntariness. Regarding automaticity: As Hakli (2006) reasons: beliefs are paradigmatically formed in an automatic and involuntary manner (consider your belief ‘There’s a knock at the door!’ which you form automatically after hearing a knock at the door), while whatever mental states groups are capable of hosting are not automatic, but rather, the result of careful deliberation. And, regarding voluntariness: beliefs seem to be *essentially* involuntary; voluntary affirmations are ‘make beliefs’, not genuine beliefs. But, group inflationary ‘belief’ comes apart from ordinary individual-level belief in both of these ways: group ‘beliefs’ are never automatic, and always voluntary. So, we should reject that group ‘beliefs’ are beliefs.

The above line of argument against group inflationary beliefs unhelpfully runs together a number of things that should be kept apart. By being more careful here, we can make some concessions to the rejectionist about group belief while upholding a version of group belief inflationism, one on which we can unproblematically attribute judgements to groups.

A first concession to the rejectionist is that anything that is a belief (whether the host of the belief be an individual or a group) is essentially non-voluntary. A second concession is that at least some individual beliefs are clearly automatic, and, *if* there are group inflationary beliefs, such group beliefs would never be automatic.²⁴ What follows from these concessions, about group beliefs, is just that: groups can't have beliefs that are either voluntary or automatic.

Does the postulation of a group *judgement* violate either of these conditions we're conceding to the rejectionist? The answer, I want to suggest, is 'no'. Just consider the following minimal statement of what a group must *do* in order to make a judgement, as construed within a telic virtue epistemology:

(Non-summative) collective judgment: A group *G* judges that $\langle p \rangle$ if and only if the *G* constitutively attempts, with intention, to get it right (whether $\langle p \rangle$) aptly by alethically affirming that $\langle p \rangle$.

Such a collective judgement will not be 'automatic'; it is (like an individual judgement) an intentional action. But does it *matter* that we've conceded to the rejectionist that *no* collective beliefs can be, like at least some individual beliefs, automatic? No; a collective being unable to produce automatic beliefs, as individuals can, is of course compatible with it being able to make judgements. For a thought experiment: imagine an individual who, operated on by neuroscientists, was unable to form beliefs in any other way than by intentional judgement. This individual would be unusual, disadvantaged even. But this inability does not call into doubt the individuals' capacity to make judgements, not in the least. Such an individual would therefore hardly be incapable of having beliefs, and even knowledge (whenever these judgements are apt), and *mutatis mutandis*, for groups—analogously unable to believe automatically while capable of judging.

Likewise, the concession to the rejectionist that beliefs can't be voluntary is not problematic for the prospects of group judgement. It would be if and only if group judgements are voluntary in virtue of being intentional. But the kind of intention characteristic of judgement at the individual level was already shown not to imply a kind of voluntariness incompatible with any sort of knowledge-apt belief. The same holds at the collective level, when a group (collectively) makes a constitutive attempt, with intention, to get it right aptly by alethically affirming. Granted, a group could very well—for example, by jointly committing

to a proposition—make a (collective) attempt to get it right (as a group) in the endeavour to do something *else*—for example, to let a proposition stand as the group’s view if and only if doing so would be strategically wise from a marketing perspective. (For example, imagine Philip Morris’ board jointly committing to the proposition that deaths from cancer are unfortunate.)²⁵ Even when such propositions are true—and indeed, even if the individuals of the group know this proposition to be true, individually—the collective attitude would not be a judgement, as the collective aim is something other than getting it right aptly. Likewise—and analogously at the individual level—a group could (if competing in a game show) jointly commit to an arbitrary answer to a game show question, and in doing so, jointly accept a proposition by merely (jointly) alethically affirming the proposition, though without judging.

Let’s briefly take stock. What the foregoing shows is that the basic ‘template’ for an individual judgement, on telic virtue epistemology, can be transferred to the collective level as a view of (inflationary) group judgement without falling foul of the kinds of objections that are so often raised to so-called group beliefs. And this is the case even if these objections (or some version of them) are applicable to some *non-judgemental* accounts of inflationary group belief.

Moreover, the basic view—that a group judgement is a constitutive attempt, with intention, to get it right aptly via alethic affirmation—is, conveniently, compatible with different views of the metaphysical nature of group belief, that is, different views about what kind of mechanisms have to take place among group members to materially realise a group belief. Let’s briefly now consider how the template proposal could potentially be glossed differently on two leading proposals types: (i) Margaret Gilbert’s (1987, 2013) *joint commitment account*; and (ii) *social-distributed accounts*, a Durkheimian functionalist version of which is defended by Alexander Bird (2010, 2019) and a cognitive integrationist version of which is defended by S. Orestis Palermos (2016).

Consider first Gilbert’s *joint commitment* model. On this view, it is necessary and sufficient for a group to believe a proposition, $\langle p \rangle$, that the group *jointly accepts* that $\langle p \rangle$. Further, the members of a group jointly accept that $\langle p \rangle$ when the members conditionally commit to accept that $\langle p \rangle$, which they do if and only if each is committed to acting as if $\langle p \rangle$ provided the others do. Thus, on this view, it is the obtaining of this joint commitment, by individual members of the group, to act as if $\langle p \rangle$ is true provided others do, that gives rise to, or ‘realises’, the group belief.

Here is how a non-summativist account of collective judgement could be given a metaphysical gloss along the lines of the above Gilbert-style account, according to which joint commitments are the realisers of group belief. First, we assimilate a group’s joint acceptance that $\langle p \rangle$ with (mere) affirmation that $\langle p \rangle$, such that a group affirms that $\langle p \rangle$ if

and only if the group members jointly accept that $\langle p \rangle$. An affirmation is an *alethic affirmation* if and only if it is an attempt to get it right (on whether $\langle p \rangle$) that $\langle p \rangle$. Now, if the individuals in a group privately desire that the group jointly accept that $\langle p \rangle$ just in case by doing so they would be (jointly) affirming truly, would *this* suffice to make the kind of attempt the group is making (when jointly committing that $\langle p \rangle$) an *alethic* affirmation, and not only an affirmation?

The answer to this question is ‘no’. The best way to interpret the above scenario is as one of a *mere* collective affirmation, given that the group itself is not affirming in any endeavour (for truth, aptness or anything else) *as* a group, even if individuals are so endeavouring when jointly accepting (and in doing so, affirming) the proposition as a group member. Rather, a group *alethically affirms* a proposition only if the group has already *collectively* established (viz., through a prior joint commitment) the scope of their endeavour to be an attempt to get it right (on whether $\langle p \rangle$). With this idea in hand, we can now characterise a group *judgement* (on the joint acceptance model) in terms of a group alethic affirmation. To a first approximation, a group judges that $\langle p \rangle$ only if, antecedent to affirming whether $\langle p \rangle$, the group jointly commits to (i) *alethically* affirm whether $\langle p \rangle$; (ii) to get it right $\langle p \rangle$ *aptly* through (i). The combination of (i) and (ii) establish that the kind of attempt at getting it right the group makes when jointly accepting (i.e., affirming) that $\langle p \rangle$ lines up with collective judgement, rather than, say, with *mere* collective affirmation or *mere* collective alethic affirmation.

Our (non-summative) collective judgement template can just as easily be given a very different gloss, if paired with a different view about what realises group belief. Consider, for example, *social distributed* views of group belief, according to which cognition involves relatively tightly integrated groups working together, with scientific research teams being the classic example (e.g., Bird 2010; De Ridder 2014; Palermos 2016). On these views, it is in virtue of the social relations at work between group members that different parts of the system contribute to the generation of the system’s collective mental state that $\langle p \rangle$, and even if (though not only if) no individual in the group actually hosts the belief that $\langle p \rangle$.²⁶ For example, suppose a scientific research team—inquiring into whether there are over 1000 species of a certain kind of bird, *B*—divides up tasks with the plan of having each individual input their own data (on the basis of performing certain individual epistemic tasks corresponding with their roles in the group) in a centralised database, which then combines the data, spitting out a collective result in the affirmative if and only if the aggregated data compiled through the shared database identifies over 1000 species of *B*. On this kind of view, the *group* takes a representational stance on the matter of whether $\langle p \rangle$ (whether the number is above or below 1000)—a belief according to social distributed views. And while the individual members may permissibly also hold the group

belief at some point in the process (e.g., by consulting the database, conferring with each other, etc.) their doing so isn't what realises the group belief. Rather, it's the distributed contributions of the individuals to the collective result in accordance with their roles that is doing the work.

One internal dispute among proponents of social-distributed models of group belief concerns the matter of how to delineate who exactly should 'count' as part of the group that has the belief it has in virtue of the distributed epistemic contributions of its members. Consider—to use a case discussed by Simion, Carter, and Kelp (2020)²⁷—the mailperson, whose job it is to bring the mail to the scientific research team—including some important research documents—and in doing so, causally contributes to the research team's group belief $\langle p \rangle$. Is the mailperson thereby a *member* of the group that believes that p ?²⁸ Intuitively, it seems the answer here should be 'no', despite the mailperson's making an epistemic contribution to the group belief in light of occupying a social role. Whereas a social-role functionalist account like Bird's has a difficult time explaining why the mailperson should be ruled out, S. Orestis Palermos's (e.g., 2016) cognitive integrationist version of a social-distributed account of group belief can do so easily. For Palermos, the relevant social interactions between group members that serve to collectively realise, via distributed cognitive tasks, a group belief, must include *feedback loops*—viz., two-way causal interactions with other contributing group members, of the sort that are, on dynamical systems theory (DST), the mark of dynamical systems. Because the causation in the case of the mailperson is asymmetrical (from the mailperson to the research team, but not vice versa), the mailperson is not a part of the group that believes despite the causal epistemic contribution made.

That said, the feedback loop requirement itself may be too strong, as it's not obvious that all *bona fide* members of, for example, a scientific research team that produces a result must interact via feedback loops with each other. For our purposes, we remain neutral about how a social-distributed model might 'thread the needle' to accommodate the above kind of dilemma. Rather, it suffices to register that an account of collective judgement could be glossed on a social distributed model, regardless of whether the relevant social relations are social-functional contributions (Bird) or reciprocal-causal contributions (Palermos). To a first approximation, the idea is as follows: A group G judges that $\langle p \rangle$ if and only if the G constitutively attempts, through *distributed* individual contributions to the group attempt, to get it right (whether $\langle p \rangle$) aptly by alethically affirming that $\langle p \rangle$.

In order to make this idea concrete, it will be helpful to compare a group judgement, glossed on the social-distributed model, with a *mere* group affirmation, also glossed on the model. For ease of reference, let's suppose we have two scientific research teams, 'Team Affirmation' and 'Team Judgement', each aiming to establish—to continue here with our

previous example of distributed cognition—whether there are over 1000 species of bird *B*. Team Affirmation, low on research funding, is trying to cut corners. The interactions between group members reflect this—none of the individuals are accountable for the accuracy of their results (as long as they register a plausible enough looking number indicating how many species they have found of bird *B* in their designated sector). With this kind of epistemic laxity characterising the social norms in play, the group eventually affirms a collective answer (fewer than 1000 species)—and suppose even that this is true. Even so, we have here a mere (collective) affirmation, and (depending on how the details are further filled out) at the very most, an *alethic* affirmation. But Team Affirmation falls short of making a judgement.

Team Judgement—true to their namesake—approaches things differently. The social norms governing the individual contributions to the collective output, as well as the interactions between individuals in ensuring reliability in reporting and collating the individual epistemic contributions, are knowledge directed. Suppose even that these norms are deeply internalised as well as explicit in the research team's manifesto: individual contributions to the group output are accepted only if reliable methods are used, and individual epistemic contributions are even cross-checked by other team members as a matter of policy to minimise epistemic risk. Suppose finally that the result is the same (fewer than 1000 species).

Given that both of our two research teams generate, through organised distributed efforts, a collective representational output (i.e., that there are fewer than 1000 species of bird *B*), both (on a social-distributed model) count as at least affirming this, perhaps even as both alethically affirming this. But only Team Judgement is actually aiming at *aptness*, and not merely at correctness, in a way that is characteristic of judgement. Team Judgement is affirming with the aim of (through the distributed intellectual contributions) getting it right *aptly*, viz., through not just any kind of way of organising and aggregating individual contributions, but—as the social norms governing their inquiry prescribe—through a reliable, knowledge-conducive way of doing so.

6 Collective Judgemental Knowledge

Let's take stock. We've seen how judgemental belief can be realised in a promising way at the collective level such that it is structurally analogous, on a telic theory of epistemic normativity, to how it is realised at the individual level—viz., through a (collective) intentional attempt to get it right *aptly* (whether $\langle p \rangle$) by alethically affirming that $\langle p \rangle$. Moreover, we've seen that an advantage of the proposal is that it is in principle compatible with competing views—viz., joint acceptance accounts and social-distributive accounts—of how group members must interact in order to materially realise a group belief.

But two residual questions remain.

Question 1: First, if a collective judgement is to result in collective judgemental knowledge (alternatively: in knowledge full well), the collective judgement must itself be apt. What, then, is required in order for a collective—as opposed to merely an individual—judgement to attain the status of aptness?

Question 2: How much collective judgemental knowledge is there? If none, or very little, then the proposed view (paired with a $K = AB$ view at the individual level) secures substantive symmetry, but not full-blown non-sceptical substantive symmetry of sort desired.

Let's now answer these two questions in turn.

(a) *An answer to Question 1*

Any judgement—just like any (constitutive) attempt, more generally—is apt if and only if it is successful and the success is through *competence*. A ϕ – competence, generally speaking, is a disposition of a subject to succeed reliably enough, whenever one makes a ϕ – attempt and is in proper shape and properly situated. (In the case of performing a triple axel, for example: if you were unable to succeed when you attempt to perform the jump while drugged and strapped inside an airplane—away from an ice rink—this would *not* count against your competence to land a triple axel. What matters is whether you'd succeed reliably enough if you tried while in proper shape (undrugged) and properly situated for such an attempt (equipped with skates, on an ice rink, plenty of ambient oxygen, etc.).

'Successful' judgement is equivalent to *apt alethic affirmation*—viz., that which judgement, as such, is a constitutive attempt to bring about. And so: a *competent judgement* is a judgement that manifests a disposition (on the part of the judging subject) to (reliably enough) succeed at doing *that*—viz., that which constitutes successful judgement. An individual thinker, for example, would possess such a competence only if disposed to reliably enough affirm alethically only if she would do so aptly. In light of such competence, the thinker *not easily* would fail to affirm with alethic aptness.

In sum, then: a judgement is *apt* just in case its success, that is, its securing the aim of apt alethic affirmation manifests (or: is because of) one's disposition to (reliably enough) affirm alethically only if one would do so aptly. We can now—drawing from our template account of collective judgement from §5—extend this account of apt judgement to the special case of where the judgement is a collective judgement as follows: (Non-summativist) **collective apt judgment (i.e., collective judgemental knowledge):** A group G judges that $\langle p \rangle$ aptly if and only if G

constitutively attempts, intentionally, to get it right (whether $\langle p \rangle$) aptly by alethically affirming that $\langle p \rangle$; (ii) G secures this aim; and (iii) G's securing this aim manifests (or: is because of) G's disposition to (reliably enough) collectively affirm alethically whether $\langle p \rangle$ only if G would do so aptly.

Important here are two key points, the first of which has to do with clause (iii)—viz., the idea that G must have a disposition—one that it manifests in judging successfully—to (reliably enough) collectively affirm alethically only if G would do so aptly. This disposition, it should be emphasised, is a disposition of the judging *group*, one the group has in an inflationary sense, such that it's false that the group possesses the disposition if and only if individuals in the group possess this disposition.

The second idea concerns the 'because of' locution. Even if a group judgement is both *successful* (i.e., it results in apt alethic affirmation) and *competent*, it might still fall short of being *apt, qua* judgement, if the group's success is not *because* competent. Suppose, for example, that a research team *competently* judges that $\langle p \rangle$, and further, that $\langle p \rangle$ is true. However, here's the twist: due to a fluke error in some of the computer equipment that the group uses to take the measure of its members' individual contributions to the group judgement (whether it be a joint-commitment-registering machine, in a Gilbert-style model, or a data aggregating machine, on a social-distributed model), the machine first (i) incorrectly registers the group's view as not- $\langle p \rangle$; but, then, (ii) a second computer glitch occurs, causing the machine to issue an arbitrary result that happens to be $\langle p \rangle$. In this case, the group lacks judgemental knowledge, (and even apt alethic affirmation) as the fact that the group affirms what it does rather than something else is not due to any kind of competence but just to dumb luck.

(b) *An answer to Question 2*

In answering Question 2—about whether group judgemental knowledge is, on the proposal advanced here, about as common as we'd expect—I want to draw a brief parallel to the individual level. Is *individual* judgemental knowledge rare? Why would we think it is? Three kinds of arguments that might try to establish this (none of which is very persuasive) go as follows: group judgemental knowledge would plausibly be rare, on the telic theory proposed, if (i) *all* knowledge—judgemental or otherwise—is rare due to the epistemic hostility of our environment (e.g., if we are being often deceived); (ii) if there are barriers to individuals performing the act of judgement given what this involves on telic virtue epistemology; (iii) if judgement itself (as it is understood on telic virtue epistemology) is—despite there being no barriers to performing it—rarely performed nonetheless.

I'm going to set aside (i) out of hand, as it has very little bearing on what's of interest here, given that its implications concern much more than just judgemental knowledge. That said, (ii) and (iii) concern judgemental knowledge more directly. Regarding (ii): We've seen in §4 that typical arguments against the possibility of voluntary beliefs cut no ice against the idea that judgements are a species of intentional action. However, with this point established, it's unclear why there should be any barrier to individual judgement, or (iii) for that matter, why—given our interest in getting it right knowledgeably (rather than just getting it right anyway) when we inquire—judgement is something an individual would rarely perform.

The point of working through the implausibility of any of (i–iii) as reasons to doubt that judgemental knowledge at the individual level is about as widespread as we'd expect is that *the same* rationale extends—*mutatis mutandis*—to the collective level. Again, setting aside (i), let's consider (ii). §5 showed in some detail why typical objections to the countenancing of inflationary group beliefs do not carry over to (inflationary) group judgement, at least as it is construed within a telic framework. Regarding (iii): there is no reason at all to think that such collective judgements are rare. On the contrary: collective judgement (as opposed to collective *mere* alethic affirmation) plausibly best describes what scientific research teams often do when endeavouring not merely to get it right any old way on a particular inquiry, but to get it right knowledgeably, and on this point, juries are no different.²⁹

7 A Comparison

There is one other attempt in the literature to extend the (K = AB) model from individual to collective epistemology, due to Jesper Kallestrup (2020), in his paper 'Group Virtue Epistemology'. There is a lot to like about Kallestrup's project. For one thing, Kallestrup maintains that there is 'nothing in [Sosa's] framework precludes ascriptions of knowledge to group agents'. Agreed. Even more, Kallestrup accepts inflationary (i.e., irreducibly collective) epistemic properties in his concession that it is 'perfectly possible for groups to instantiate epistemic properties none of their members instantiate'. Also agreed. Thirdly, Kallestrup thinks that the right way to think of group knowledge is as apt belief, and even more, his position allows that a 'group may form apt beliefs none of its members share'. This third point, along with the other two, line up exactly with my own thinking.

But Kallestrup and I—despite travelling the same road, from individual to collective epistemology, part of the way—diverge at several important places. I want to conclude by showing why I think that the view defended here, and not Kallestrup's, gets it right on these key points of divergence.

(a) *First point of divergence*

According to Kallestrup, while a group may form apt beliefs that none of its members share, the ‘competence’ of a group is nothing over and above the competences of its members when suitably combined. He writes:

Novel competences of groups do not spring into existence or mysteriously emerge when conjoining existing individual ones. On the other hand, the aptness of group belief is not similarly reducible to the aptness of the beliefs of its individual members.

(2020, 5242)

In a bit more detail, Kallestrup’s contention is that the following thesis is true of group competences but false of knowledge *qua* apt belief

Reductive individualism: all (or at least most) of the members of *g* having E-type properties is necessary and sufficient for *g* having E.

(2020, 5247)

There are two problems with this view. Firstly, empirical evidence for cases of ‘Mandevillian intelligence’ indicates that reductive individualism does not hold for group competences.

Cases of Mandevillian intelligence, in short, indicate that such a reduction is problematic because some dispositions that are *unreliable* and thereby are not individual-level competences can, and reasonably often enough do, lead to knowledge-conducive dispositions at the collective level, particularly when these individual-level shortcomings play the *de facto* role of generating cognitive diversity within a group. In a series of recent papers reviewing these kinds of cases, Paul Smart (2018a) has noted individual-level ignorance, extreme-thinking and forgetfulness as (somewhat paradoxically) among the individual-level traits that have been reported as contributing to epistemic benefits at the group level. Whereas the countenancing of Mandevillian intelligence cases is incompatible with Kallestrup’s reductive individualism about group-level competences, it is compatible with my proposal which, unlike Kallestrup’s, does not make this commitment.

A second problem for Kallestrup’s reductive individualism about group competence is that it is paired with *non-reductive individualism* (i.e., the denial of reductive individualism) about group knowledge *qua* apt belief. This pairing is unstable in the following way: it posits a non-reductive subject as possessing an apt belief while denying that the *same subject*, the subject to whom knowledge is attributable, can possess a competence. This pairing implies that Kallestrup must reject the

following which is analytically true within the kind of virtue-theoretic framework he takes himself to be operating: a subject, *S*, has an apt belief only if *S* has a competent belief.

(b) *Second point of divergence*

The second point of divergence concerns theoretical neutrality. According to (K = AB) proposals, knowledge is a *normative (epistemic) kind*, not a psychological kind. Such theories, at the individual level, do not carry any heavy-duty commitments about how beliefs are materially realised in the mind of a knower. For this reason, a proponent of (K = AB) at the individual level could potentially be paired with various kinds of views about beliefs in the philosophy of mind, for example, from representationalism to functionalism. The version of the (K = AB) template opted for here (§5) was shown to be compatible with very different theories of what has to happen to realise a knowledge-apt collective judgement. In this respect, the proposal is, like individual-level (K = AB) views, theoretically neutral about the conditions on realising belief. Kallestrup's proposal, by contrast, weds itself to several heavy-duty commitments on this front. To give one such example, for Kallestrup, an (inflationary) group belief is brought about, through the activity of individual members, *only* if those individuals' contributions are made (i.e., only if the individuals form the beliefs they form) *because* they believe that others intend to make the certain prescribed individual contributions towards the group attitudes (2020, 13). This caveat alone will make the view off-limits to social-distributed views that lack any such explicit requirement. Even more, the view seems to generate some implausible predictions. For example, the view predicts that a scientific research team fails to generate group belief (and thus group knowledge) if each individual on the research team, in short, *would have kept on working* exactly as they had even had the others bunked off. And this looks like the wrong result.

(c) *Third point of divergence*

The third point of divergence concerns the matter of 'screening off' rejectionist-style objections from voluntariness. Just as individual-level acceptance is voluntary in a way that individual-level knowledge-apt belief is not, likewise—and this is just a restatement of the kind of point that has been made by various rejectionists about group belief (see §5)—a voluntary acceptance (joint or otherwise) of a proposition is not a knowledge-apt belief.

I've shown how collective *judgements*, though intentional, are not just usually, but essentially, not voluntary, and this result falls out of the normative structure of judgement as a distinctive kind of constitutive attempt, as understood within telic virtue epistemology.

Kallestrup's proposal, at the collective level, does not (as mine does) unpack the (K = AB) template as a thesis about collective *judgemental knowledge* specifically—knowledge that implicates an intentional (and successful) constitutive aiming at the aptness of alethic affirmation rather than at any other end—but rather, as a thesis about collective knowledge more generally, knowledge generated whenever collective 'belief' is apt. But here is the problem: there is nothing in this proposal that screens off the objection that group 'beliefs' aren't genuine (knowledge-apt) beliefs, given that—in light of how Kallestrup has described their conditions of realisation—they could in principle be brought about voluntarily. After all, on Kallestrup's proposal, group beliefs—even though they purport to describe the world and as such are attitudes with a mind-to-world direction of fit—can be generated and reversed arbitrarily by a group so long as the group satisfies (following List and Pettit (2006)) conditions whereby the group acts intentionally to have that attitude stand *as* the group's attitude (see, e.g., 2020, 5241). But then, and taking a wider view, susceptibility to this kind of criticism leaves Kallestrup's version of collective virtue epistemology one that preserves substantive symmetry (see §2) only at the cost of inviting the charge that the view is not suitably *non-sceptical* at the collective level, even if substantively symmetrical.

8 Concluding Remarks

In sum, the simple idea, powerful in individual epistemology, that knowledge = apt belief is not a 'one-trick pony'. It can be made to work at the collective level as well. But in order to make it work, we have to understand *how* to make it work, and that requires availing ourselves of the full theory of telic epistemic normativity, within which we can distinguish different kinds of beliefs in light of the distinctive kinds of constitutive attempts they are at getting it right. With this in mind, the view is that 'K=AB' is the correct theory of knowledge at the collective level *as a theory of collective judgemental knowledge*, or of *apt judgement*. Understood as such, the view was shown to be insulated against standard fare criticisms of collective belief that would seem, *prima facie* to pose problems for any account of collective knowledge that is 'built' out of any sort of collective belief. Moreover, the view is shown to diverge from an alternative view of collective virtue epistemology defended by Jesper Kallestrup in three important ways—and I've argued why, at each of these three forks in the road, there are key advantages to throwing in with the theory defended here.

Acknowledgements

Thanks also to Mark Alfano, Jeroen de Ridder and Kate Devitt for helpful comments on a previous version of this paper.

Notes

- 1 For a notable presentation of this idea, see Sosa (2007).
- 2 See, for example, Sosa (2010a).
- 3 See, in particular, Sosa (1997a, 2009, Ch. 2). Cf., Carter (2020).
- 4 See Sosa (1997b).
- 5 E.g., Sosa (2010b).
- 6 See here, in particular, Sosa (2020, Chs. 3–6).
- 7 One exception is Kallestrup (2016), whose work will be discussed here.
- 8 See Sosa (2020, 20–23).
- 9 Though, for some kinds of beliefs (e.g., judgments), not merely at truth. This point is taken up in §4.
- 10 For an overview, see Simion, Carter, and Kelp (2020).
- 11 Note that I'm not considering all possible combinations here that pair ($K = AB$) with a position type at the collective level.
- 12 For some representative discussion on this point, see, for example, Tollefsen (2015).
- 13 See Simion, Carter, and Kelp (2020).
- 14 Cf., Williamson (2000).
- 15 For examples of the former, see, e.g., Bird (2010), Palermos (2020), and De Ridder (2014). For an example of the latter, see Hakli (2006).
- 16 See, e.g., Kelp (2018) and Simion (2019).
- 17 See Sosa (2015, Ch. 3).
- 18 For an extended discussion on this point, see Sosa (2015, Ch. 3, fn. 5).
- 19 See Sosa (2020, 24–26).
- 20 See Sosa (2020, 29).
- 21 Sosa (2015, 166).
- 22 See, for discussion, Sosa (2007, 89) and especially Sosa (2020, Ch. 2).
- 23 See Gilbert (1992, 257–258); cf., Bird (2019, 3).
- 24 To be clear, I'm just granting this here to the rejectionist for the sake of argument.
- 25 Cf., Lackey (2020).
- 26 For discussion, see Simion, Carter, and Kelp (2020, sec. 3).
- 27 The original example of the mailperson is due to Mona Simion.
- 28 As Mark Alfano has pointed out to me, a variation on this kind of philosophical issue is on display in cases of adjudicating CERN authorship. See <http://library.cern/cern-author-guidelines>.
- 29 See Bird (2007a, 2007b).

References

- Bird, Alexander. 2007a. "Justified Judging." *Philosophy and Phenomenological Research* 74 (1): 81–110.
- . 2007b. "What Is Scientific Progress?" *Noûs* 41 (1): 64–89.
- . 2010. "Social Knowing: The Social Sense of 'Scientific Knowledge'." *Philosophical Perspectives* 24: 23–56.
- . 2019. "Group Belief." In *The Routledge Handbook of Social Epistemology*, edited by Miranda Fricker, Peter J. Graham, David Henderson, and Nikolaj Pedersen. Routledge: 274–283.
- Carter, J. Adam. 2020. "Epistemic Perceptualism, Skill and the Regress Problem." *Philosophical Studies* 177 (5): 1229–1254.
- De Ridder, Jeroen. 2014. "Epistemic Dependence and Collective Scientific Knowledge." *Synthese* 191 (1): 37–53.

- Gilbert, Margaret. 1987. "Modelling Collective Belief." *Synthese* 73 (1): 185–204.
- . 1992. *On Social Facts*. Princeton University Press.
- . 2013. *Joint Commitment: How We Make the Social World*. Oxford University Press.
- Hakli, Raul. 2006. "Group Beliefs and the Distinction between Belief and Acceptance." *Cognitive Systems Research* 7 (2–3): 286–297.
- Kallestrup, Jesper. 2020. "Group Virtue Epistemology." *Synthese*, 197: 5233–5251.
- Kelp, Christoph. 2018. *Good Thinking: A Knowledge First Virtue Epistemology*. Routledge.
- Lackey, I. Jennifer. 2020. "Group Belief: Lessons from Lies and Bullshit." *Aristotelian Society Supplementary Volume*, 94: 185–208. Oxford University Press.
- List, Christian, and Philip Pettit. 2006. "Group Agency and Supervenience." *The Southern Journal of Philosophy* 44 (S1): 85–105.
- Palermos, S. Orestis. 2016. "The Dynamics of Group Cognition." *Minds and Machines* 26 (4): 409–440.
- . 2020. "Epistemic Collaborations: Distributed Cognition and Virtue Reliabilism." *Erkenntnis*, 1–20. <https://doi.org/10.1007/s10670-020-00258-9>
- Simion, Mona. 2019. "Knowledge-First Functionalism." *Philosophical Issues* 29 (1): 254–267.
- Simion, Mona, J. Adam Carter, and Christoph Kelp. 2020. "On Behalf of Knowledge First Collective Epistemology." In *Doxastic and Propositional Warrant*, edited by Paul Silva and Luis Oliveira. Routledge.
- Smart, Paul R. 2018a. "Mandevillian Intelligence." *Synthese* 195 (9): 4169–4200.
- Sosa, Ernest. 1997a. "Mythology of the Given." *History of Philosophy Quarterly* 14 (3): 275–286.
- . 1997b. "Reflective knowledge in the best circles." *The Journal of Philosophy*, 94 (8): 410–430.
- . 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume 1*. Oxford University Press.
- . 2009. *Apt Belief and Reflective Knowledge Volume II: Reflective Knowledge*. Oxford: Oxford University Press.
- . 2010. "How Competence Matters in Epistemology." *Philosophical Perspectives* 24 (1): 465–475.
- . 2015. *Judgment and Agency*. Oxford University Press.
- . 2020. *Epistemic Explanations: A Theory of Telic Normativity, and What It Explains*.
- Tollefsen, Deborah Perron. 2015. *Groups as Agents*. John Wiley & Sons.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford University Press.

11b Commentary from Jeroen de Ridder

Achieving Collective (Telic) Virtue-Theoretic Knowledge

Adam Carter's essay puts yet another feather in the cap of Ernest Sosa's virtue epistemology. Carter shows how Sosa's central proposal of construing knowledge as apt belief provides the resources for an account of collective knowledge that has the following desirable features: (1) It is non-reductive or non-summative, which is to say that it makes collective knowledge a robustly group-level phenomenon that is irreducible to a mere sum of individual knowledge. (2) It is non-sceptical; some groups really can and do possess collective knowledge. And (3) it is pleasingly theoretically unified, offering a single virtue-theoretic framework to make sense of both individual and collective knowledge.

Carter presents his proposal for a virtue-theoretic account of collective knowledge at a high level of abstraction, which shines a helpful light on the structural parallels between individual and collective knowledge. But here, I want to explore what happens when we put the proposal to work in a concrete potential case of collective knowledge, to get a better sense of when groups know. Recall what is required:

a group judges that $\langle p \rangle$ only if, antecedent to affirming whether $\langle p \rangle$, the group jointly commits to (i) alethically affirm whether $\langle p \rangle$; (ii) to get it right [whether] $\langle p \rangle$ aptly through (i).

When the group's attempt to judge collectively in this manner is successful, it knows (full well) that p . Let's take a familiar example from the literature on the joint commitment account of collective belief—a job search committee—to see what this entails.

We'll assume hospitable conditions for knowledge acquisition: the committee consists of fair-minded academics who have no personal axes to grind and who, through their work on the committee, intend to select the candidate who is the best match for the department, all things considered. Let's also assume that who the best candidate is, is not a matter of taste; there is an objective matter of fact about who it is, so

that knowledge is in principle possible. To satisfy (i), all (or most) committee members—or at least the operative ones (cf. Tuomela 2004) — must commit to making an alethic affirmation as a group about who the best candidate is, conditional on it being common knowledge that others make the same commitment. To satisfy (ii), members must also commit to the use of a decision procedure, which will determine what their alethic affirmation as a group will be—for instance, consensus deliberation, voting, or expert consultation. This commitment, too, is conditional on it being common knowledge that other members are similarly committed. Since the goal is to *get it right aptly* through alethic affirmation, the procedure has to satisfy further conditions. Committee members must take it to be truth-conducive, which is to say they (implicitly) believe that its proper use is likely to produce the correct view about who the best candidate is. Proper use, in turn, might mean different things: either committee members can think that the procedure is generally reliable to settle a question like this when properly executed, no matter the quality of the inputs; or they can think that the procedure is reliable on this particular occasion, given how it is in fact executed and given the quality of the input that they collectively provide.

To summarise, in order for a group to acquire virtue-theoretic knowledge that p on the joint commitment model, group members must:

- Commit to alethically affirm that p as a group;
- Know that all other group members are similarly committed to alethically affirm that p as a group;
- Commit to the use of a decision procedure for determining their view as a group;
- Know that all other group members are similarly committed to the use of this procedure as a group;
- Believe (perhaps implicitly) that this procedure is truth-conducive, either in general or at least on this particular occasion, which is to say they must believe (perhaps implicitly) that their use of the procedure makes it likely to produce a true output.

This list makes you wonder how much group knowledge (thus construed) there will be: how often do groups satisfy all of the above conditions? Will members of a job search committee really think that deliberating with their colleagues or taking a vote is likely to produce the correct view about the best candidate? If one committee member suspects another of strategic voting or incorrectly weighing the evidence, she will have doubts about the truth conduciveness of the procedure and the commitment breaks down. Or if one member harbours doubt about whether deliberation has been fair and reasonable, she might question the reliability of the procedure on this occasion. Similarly, if just one

member feels they ought to take a vote rather than deliberate, or defer to a colleague who has relevant expertise, the joint commitment to the decision procedure breaks down.

While search committees and similar task-oriented groups are at least likely to agree on a procedure before they start their work, more organic groups such as teams of researchers often don't (Bird 2010). In so far as they even end up accepting some view as the team's official view, it's doubtful whether all members will really be jointly committed to thinking that the process they happened to use on that particular occasion is truth conducive. Remember that even one dissenting opinion or doubt can be enough to undermine the joint commitment.

The upshot is that the joint commitment version of virtue-theoretic collective knowledge may be quite rarely instantiated. Much rarer, at any rate, than colloquial ascriptions of group knowledge would suggest. Collective telic virtue epistemology may be more sceptical and revisionist than it appears, even if Carter is right that it occupies an attractive position in theoretical space.

References

- Bird, Alexander. 2010. "Social Knowing: The Social Sense of 'Scientific Knowledge.'" *Philosophical Perspectives* 24 (1): 23–56. <https://doi.org/10.1111/j.1520-8583.2010.00184.x>.
- Tuomela, Raimo. 2004. "Group Knowledge Analyzed." *Episteme* 1 (2): 109–127. <https://doi.org/10.3366/epi.2004.1.2.109>.

11c Commentary from Kate Devitt

Turtles All the Way Down: Automaticity, Involuntariness and the Symmetry of Group Beliefs

On 9 April 2021, Prince Phillip died. The announcement of this death was disseminated almost instantly over traditional media and social media channels and involuntarily created a collective true belief. Many individuals formed the belief via well-honed and competent testimonial processes, for example by reading a trusted news source, such as ABC News (Australia) or BBC news (globally). But, more importantly (for this volume), human society as a whole formed this belief automatically and involuntarily regardless of the fact that billions of individual humans across the earth held no such belief. This is a success story for an inflationist (not summative), animal knowledge account of group belief. This is because the global knowledge networks of humanity have been preparing for this announcement for years so that the global announcement of Phillip's passing was fast and apt. The doctors informed the palace, the palace released a statement, and the global media communicated the sad news collectively and in parallel; each newspaper, organisation or outlet releasing their own stories to complement the singular fact of the matter. Groups can have knowledge and it does not need to come from careful deliberation at the moment of knowledge acquisition. Society heard 'the knock at the door' that Prince Phillip had died and knew that it was so. Like the golfer who practises holes-in-one or the basketball player who trains their skills to shoot hoops, groups practise information gathering, processing and disseminating in advance of significant and predictable real events in the world so that they perform aptly on the day. Human society collectively knowing that Prince Phillip has died is a constitutive attempt, after years of instrumental practise receiving propositions of varying veracity regarding the royals.

If the account above is true, then it means that the disanalogies claimed between individual belief on the one hand and inflationary group belief on the other, namely (i) automaticity and (ii) involuntariness, fail to stump the advocate for group belief and pave the way for a symmetric account of group knowledge. The other argument against the properties

of individual beliefs carrying into group beliefs is that beliefs are essentially conscious, but a functionalist does not require consciousness from beliefs. See also the long tradition of occurrent or explicit beliefs versus tacit beliefs (Price 1969; Lycan 1986) or aliefs (Gendler 2008). Implicit beliefs are automatic, involuntary and often hidden from conscious examination. So, three factors that make individual beliefs different from group beliefs fall away.

If symmetry is important (and I agree that it is), I'd like to introduce a model for the mind that provides a way of examining group beliefs and individual beliefs from a neuroscientific perspective. In the book 'A Thousand Brains', Hawkins (2021) describes the brain as consisting of 150,000 smaller 'brains' in cortical columns (like strings of spaghetti) through the thickness of the neocortex. Each column has a sensory-motor model of the world (forming dynamic doxastic states) and a unique frame of reference. These brain parts compare their models (aka beliefs about the world) with the models of other cortical columns and somehow vote on the most-likely-to-be-true version of the world to succeed. The unity of consciousness is achieved by the coordination of these smaller brains to form a singular belief for the human that drives actions. The unified belief is an amalgam, not a summation of 150,000 viewpoints. Neuroscience seems a fruitful way to conceive of how many human beliefs could combine into proper group beliefs. Each human is like a 'cortical column' with a model of the world and shares the output of this model with other humans to form group-level beliefs. Group beliefs are distributed over the extended mind of humanity, paper, books, phones and physical objects and landscapes. Collective epistemology would do well to explore the information exchange at both a neural and extended level, in 'embodied, emotion-rich, and environmentally modulated processes' (Gallagher 2013).

Regardless of the right account of individual and collective knowledge, I agree with Carter that Sosa's virtue epistemological framework provides a mechanism to evaluate just how knowledge-bearing individual and group beliefs are, and the category of that knowledge. Whether groups have knowledge (animal or reflective; instrumental or constitutive attempts) rather than merely justified belief, justified true belief or Gettier-proof AAA (accurate, adroit, apt) belief is particularly important for judgements of the moral responsibility of groups. Moral responsibility occurs when a situation has ethical risk, an agent has morally relevant information and the freedom to act. Agents who know the situation have more moral responsibility than those who are uncertain about the situation or those who have false beliefs. So, humanity, as a whole knows that climate change is real (regardless of deniers) and thus has a moral responsibility to act to reduce the likelihood of ethical risks. A nation that knows people are lost at sea in their territorial waters has an obligation to try and rescue them and is morally responsible for the

decisions it makes. This responsibility may be borne by many individuals of the nation such as public servants and members of the civilian government or military who have the knowledge, power, freedom and duty to act (Scott & Carr 1986). A nation that genuinely does not know what is occurring or is not able to act on knowledge is less morally responsible for harms. A corporation that does not know the environmental damage it is doing is less culpable than one that does and vice versa (see, e.g., Supran & Oreskes 2017). Sosa's framework could explain how groups gain reputation by holding beliefs with appropriate competences and justification; employing trustworthy processes for belief acquisition. Each group's review mechanisms and governance structures could be examined for their alignment to the ambitions of apt beliefs and knowing full well, as well as inspiring better ways of investing in ICT architectures and processes to improve the production of knowledge-bearing beliefs in groups.

References

- Gallagher, S. (2013). The socially extended mind. *Cognitive Systems Research*, 25, 4–12.
- Gendler, T. S. (2008). Alief and belief. *The Journal of Philosophy*, 105(10), 634–663.
- Hawkins, J. (2021). *A Thousand Brains: A New Theory of Intelligence*. New York: Basic Books.
- Lycan, W.G. (1986). Tacit belief. R.J. Bogdan (ed.) *Belief: Form, Content, and Function*. Oxford: Clarendon, pp. 61–82.
- Price, H. H. (1969). *Belief*. London: Allen & Unwin.
- Scott, G. L., & Carr, C. L. (1986). Are states moral agents? *Social Theory and Practice*, 12(1), 75–102.
- Supran, G., & Oreskes, N. (2017). Assessing ExxonMobil's climate change communications (1977–2014). *Environmental Research Letters*, 12(8), 084019.

11d J. Adam Carter's Response to Commentaries

Thanks to Jeroen de Ridder and S. Kate Devitt for their very helpful comments on my chapter 'Collective (Telic) Virtue Epistemology'. They've both given me a lot to think about, and—while I can't engage with all of their rich remarks in this brief space—I will focus on one core criticism from each and offer some thoughts in response.

I'll begin with de Ridder's comments. His critique of my proposal can be summed up simply: that it is too strong, such that it will imply that there is less group knowledge than we take there to be.

In a bit more detail, de Ridder takes issue with my characterisation of what a collective *judgement*, construed within a telic virtue epistemological framework, would demand of a group, and the worry is that it is too much. Let's look at the details. On my proposal, a group *G* *judges* that $\langle p \rangle$ if and only if the *G* constitutively attempts, with intention, to get it right (whether $\langle p \rangle$) aptly by alethically affirming that $\langle p \rangle$.¹ This core proposal is, crucially, meant to be in principle open to very different kinds of glosses in collective epistemology. It is, for example, open to a Gilbert-style² 'joint commitment' gloss as well as a distributed-cognition-style gloss. What a collective judgement would demand of its members will be different depending on whether one favours one approach rather than the other. De Ridder challenges, specifically, the shape the proposal would take *if* one were to opt to give it a joint-commitment gloss—according to which we get the result that a group judges that $\langle p \rangle$ just when antecedent to affirming whether $\langle p \rangle$, the group jointly commits to (i) alethically affirm whether $\langle p \rangle$; (ii) to get it right $\langle p \rangle$ aptly through (i).

De Ridder takes it that, suitably unpacked, judging *knowledgeably*³ will require the following of the group members: that they

- Commit to alethically affirm that *p* as a group;
- Know that all other group members are similarly committed to alethically affirm that *p* as a group;
- Commit to the use of a decision procedure for determining their view as a group;

- Know that all other group members are similarly committed to the use of this procedure as a group;
- Believe (perhaps implicitly) that this procedure is truth-conducive, either in general or at least on this particular occasion, which is to say they must believe (perhaps implicitly) that their use of the procedure makes it likely to produce a true output.

I am sympathetic to de Ridder's worry here; this does look like a lot! I'd like to canvass three lines of response. Firstly, I think we should resist the fifth of de Ridder's five proposed requirements, bearing in mind that telic virtue epistemology—on both the individual and collective level where I'm envisaging it—is externalist through and through.⁴ Second, the brunt of the requirements here are simply implicated by what joint commitment requires in simply taking up any kind of epistemic attitude.⁵ Third, and this is perhaps most important, the pairing of the core proposal with a joint-commitment account is optional; §5 shows how the view can be given different theoretical glosses when paired with a social-distributed account of group belief, including, for example, Durkheimian functionalism (Bird 2010) and dynamical systems theoretic approaches (Palermos 2020).

I turn now to Devitt's discussion, which was largely sympathetic to my proposal. For the sake of this discussion, I want to focus on one kind of alternative she considers, in the following passage:

I'd like to introduce a model for the mind that provides a way of examining group beliefs and individual beliefs from a neuroscientific perspective. In the book 'A Thousand Brains', Hawkins (2021) describes the brain has consisting of 150,000 smaller 'brains' in cortical columns (like strings of spaghetti) through the thickness of the neocortex. Each column has a sensory-motor model of the world (forming dynamic doxastic states) and a unique frame of reference. These brain parts compare their models (aka beliefs about the world) with the models of other cortical columns and somehow vote on the most-likely-to-be-true version of the world to succeed. The unity of consciousness is achieved by the coordination of these smaller brains to form a singular belief for the human that drives actions. The unified belief is an amalgam, not a summation of 150,000 viewpoints. Neuroscience seem a fruitful way to conceive of how many human beliefs could combine into proper group beliefs. Each human is a like a 'cortical column' with a model of the world and shares the output of this model with other humans to form group-level beliefs. Group beliefs are distributed over the extended mind of humanity, paper, books, phones and physical objects and landscapes.

I have two comments on the above, one ponderous and the other supportive. The ponderous comment is as follows: let's assume that the above picture is correct. If so, how would we explain a particular kind of group belief that arises only through certain kinds of normative relationships between group members. For example, consider—to borrow a case often used by Jennifer Lackey (2021)—Philip Morris's stance that there is no connection between smoking and lung cancer. How on the above proposal could we make sense of the thought that Philip Morris could hold on to this belief even when the company's individual members know better?

The supportive comment is that the above proposal strikes me as offering a potentially fruitful way to make sense of how distributed knowing—as it is developed by Edwin Hutchins (1995)—might be viewed as realised in a way that is broadly symmetrical to how individual knowledge is realised. While my chapter doesn't engage with this in much detail, an interesting line of further research would be to see just how distributed cognition, construed along the lines of an *amalgamation* as sketched above by Devitt, might be brought together with the kind of telic virtue epistemology at the collective level I've defended.

Notes

- 1 See §3 and §5 of my chapter in this volume for details of what some of the key terms here mean. For the most recent detailed account of both the notions of 'constitutive attempt' and 'alethic affirmation' as they feature in this proposal, see Sosa (2021).
- 2 See, for example, Gilbert (1987). For a more recent development on the view, see Gilbert (2013).
- 3 Within a telic virtue epistemology, a judgement (individual or collective) is apt iff its constitutive aim (viz., the aim of getting it right aptly by alethically affirming) is aptly attained. See §3 of my chapter in this volume for details; for the canonical presentation of these ideas at the individual level, see Sosa (2015).
- 4 For an early discussion of this point in bi-level virtue epistemology, see Sosa (1997).
- 5 See Mathiesen (2006) and Carter (2015) for discussion.

References

- Bird, Alexander. 2010. 'Social Knowing: The Social Sense of "Scientific Knowledge"'. *Philosophical Perspectives* 24: 23–56.
- Carter, J. Adam. 2015. 'Group Knowledge and Epistemic Defeat'. *Ergo, an Open Access Journal of Philosophy* 2. <https://doi.org/10.3998/ergo.12405314.0002.028>.
- Gilbert, Margaret. 1987. 'Modelling Collective Belief'. *Synthese* 73 (1): 185–204.
- . 2013. *Joint Commitment: How We Make the Social World*. Oxford University Press.
- Hutchins, Edwin. 1995. *Cognition in the Wild*. MIT Press.

- Lackey, Jennifer. 2021. *The Epistemology of Groups*. Oxford University Press.
- Mathiesen, Kay. 2006. 'The Epistemic Features of Group Belief'. *Episteme* 2 (3): 161–175.
- Palermos, Spyridon Orestis. 2020. 'Epistemic Collaborations: Distributed Cognition and Virtue Reliabilism'. *Erkenntnis*, 1–20. <https://doi.org/10.1007/s10670-020-00258-9>
- Sosa, Ernest. 1997. 'Reflective Knowledge in the Best Circles'. *Journal of Philosophy* 94 (8): 410–430. <https://doi.org/jphil199794827>.
- . 2015. *Judgment & Agency*. Oxford University Press.
- . 2021. *Epistemic Explanations: A Theory of Telic Normativity, and What It Explains*. Oxford University Press.

T&F Proofs – Not for Distribution

12 Three Models for Collective Intellectual Virtues

Jeroen de Ridder

1 Introduction and Preliminaries

On April 10, 2019, 3 pm, in the headquarters of the European Research Council in Brussels, the German-Dutch astrophysicist Heino Falcke, then chair of the Event Horizon Telescope science council, showed the world the first-ever image of a black hole (Devlin 2019). The image was the result of a massive collaboration involving a network of eight inter-linked radio telescopes across the globe and a team of more than 300 scientists from 60 institutes in 18 countries.¹ Falcke testifies that he originally came up with the idea for measuring a black hole's event horizon in the late 1990s, 20 years before the image was finally revealed. Clearly, the people who created this image achieved an immense epistemic success. Equally clearly, the success could not have been achieved without team effort. Collaboration was both practically and cognitively necessary: the sheer amount of work was massive and it involved a combination of scientific expertise, skills, knowledge, and understanding that no single researcher has on her own. The vernacular of intellectual virtues naturally lends itself to describing and evaluating what the team did: it displayed perseverance, creativity, curiosity, well-placed trust, and organized skepticism.

Cases like these – it's easy to multiply examples – make it plausible that the language of intellectual virtues applies as naturally to groups as it does to individuals. And hence that there are collective intellectual virtues. The purpose of this chapter is to present three models for collective virtue: three ways of understanding how collectives or groups can possess features that make them flourish and excel from the epistemic point of view.

Before I turn to these models, I should clarify a few things. The suggestion that there are collective virtues raises methodological and metaphysical questions that are familiar from the growing literature on collective epistemology (cf. Lackey 2014; Lahroodi 2019). For example: should the starting point for analyses of belief, knowledge, virtue, and other epistemic states be individuals, or should analyses remain neutral between individuals and groups (Gilbert and Pilchman 2014)? Are all ascriptions of group epistemic states shorthand for (complex) ascriptions of individual

epistemic states so that group epistemic states are reducible to sums of individual epistemic states? Do group epistemic states require group mental states or, even more problematically, group minds? What, if any, are the conditions on the individuals in a group in order for the group to be in an epistemic state or possess an epistemic quality? The worry animating these questions is that groups cannot be epistemic subjects in their own right, over and above the individuals making up the group; that groups cannot ‘really’ or ‘irreducibly’ believe, know, or possess virtues.

I’ll sidestep concerns about whether groups can ‘really’ possess intellectual virtues. Not because they are unimportant, but because I want to make progress on a different task. Namely that of understanding the different things we might mean when we use the language of intellectual virtue to evaluate group epistemic performance. I will introduce three models – or three how-possibly explanations – for how groups can have features that make them perform excellently from the epistemic point of view. For ease of exposition, I will henceforth call these features collective intellectual virtues, but, in doing so, I remain noncommittal about whether all three models ultimately describe features that are ‘really,’ or ‘irreducibly’ collective virtues.

Another issue from the literature I’ll steer clear of is what kinds of groups (if any) can possess virtues, as well as the even more fundamental one of what groups are (Ritchie 2015; Epstein 2019). The presentation of each model will make it clear what kinds and degrees of internal organization and coordinated behavior a collective must have to exemplify that model. I will leave it to others to decide whether those factors suffice for that collective to constitute a ‘real’ group or a certain kind of group, rather than a mere collection of individuals. The only constraint on my discussion is that it is focused on what I will call *collaborative* collectives or groups, that is, more or less stable groups that intentionally work together towards some common goal, such as committees, teams, organizations, departments, etc. Hence, I will not discuss arrangements in which the judgments or betting behavior of random and fleeting collections of individuals are aggregated to generate epistemically reliable outcomes.² In such cases, a collective is used by an external agent as an instrument to generate epistemically high-quality output, but does not itself form an epistemic agent in any meaningful sense.

I will also refrain from taking a stand on the nature of virtue. Some virtue epistemologists distinguish between reliabilist and responsibilist conceptions of virtue (Baehr 2006; Battaly 2008). The former conception takes reliable cognitive faculties as its model for virtues, whereas the latter takes cultivated character traits, which typically include proper motivation and emotion, as the model for intellectual virtues. I will be ecumenical here and draw on examples of both kinds, although it should be obvious that it is more difficult to argue that groups can ‘really’ have collective responsibilist virtues.

If you worry that the above qualifications undermine the motivation for thinking about collective intellectual virtue, let me point out that the task of understanding how groups can do well epistemically is relevant for collective epistemology, regardless of where the chips fall on what counts as genuinely collective virtue. Many groups in contemporary societies carry out epistemic tasks such as information gathering, storage, dissemination, or analysis, either as their primary goal (e.g., the sciences and humanities, education, journalism, R&D) or in the service of some other primary goal (e.g., administration of justice, political decision-making, governance, producing products, or offering services). To carry out such tasks successfully, collectives need to perform well epistemically and so it is important to describe and understand the different ways in which groups can do so. The general models I develop here do not straightforwardly translate into prescriptions about how to improve the epistemic life of specific individual groups or collectives – we need fine-grained empirical data from cognitive and social psychology, political science, management science, etc. for that – but they do sketch broad possibilities for how groups can flourish epistemically, which can serve as the basis for further empirical exploration and fine-tuning.

2 Addition

The first model of collective virtue is straightforward. In some cases, a group does well epistemically when all or most of its members are epistemically excellent (and nothing prevents them from exercising their individual intellectual virtues in the context of the group). A team of three creative individuals can be even more creative than the individuals working alone. Scholars who are individually intellectually perseverant can stimulate each other to become even more perseverant when they collaborate. An open-minded thinker and an intellectually generous one might make a great teaching team, and so on. The basic idea is addition: collaborating virtuous individuals exhibit virtuous behavior as a team.

This can happen in two different ways: (a) the individuals in the team might all possess the same virtue to various degrees, resulting in the collective exhibiting the dispositions and behavior relevant to that virtue to a high degree, or (b) the individuals may possess different, complementary virtues, and act accordingly in their capacities as team members, thus causing the collective to display behaviors that fit with the different virtues of the individual team members.

In collective epistemology, proposals along these lines are typically labeled *summative*. Summative models construe collective states as reducible to the mental states of the individuals who make up the collective. Saying that the group believes that *p* is thus shorthand for saying that most or all of the individuals in the group believe that *p* (cf. Quinton 1975, 17); ‘the group is intellectually humble’ means that its members

are intellectually humble. For this reason, several epistemologists hold that summative models of collective states aren't really or robustly collective (Gilbert 1987, 2014; Tuomela 1992, 2004; Bird 2010; de Ridder 2014; Lackey 2021). If talk of group belief, knowledge, understanding, and virtue is nothing more than an efficient way of talking about individuals, then there are no genuinely collective epistemic states. I won't take sides here, as I explained before. For present purposes, it suffices to note that some talk of collective intellectual virtues can indeed be analyzed as a way of saying that the individuals in the collective possess the relevant virtues.

Individually virtuous group members do not necessarily make a virtuous group on the first model. There are several reasons why individual virtues might fail to produce collective virtue. First, on the traditional Aristotelian conception of virtue, virtues are the golden mean between excess and deficiency. Collaborating individuals with the same virtue could produce an excess of the underlying trait at the group level. A group of open-minded individuals can become credulous or intellectually feeble. Second, individual virtues can cancel each other out in a collaborative setting. A creative person working together with a meticulous individual might dampen each others' individual virtues. Third, the group's formal or informal organization and culture can prohibit the manifestation of added individual virtues at the group level. A team of intellectually courageous and creative individuals might see all their sound ideas shot down in a conservative organization that overemphasizes proper procedure, due diligence, and risk avoidance.

The second and third models chart how nonvirtuous individuals can act together to produce intellectually virtuous behavior at the group level.

3 Interaction

In the second model, mutual interactions between group members and the group's structure and culture are key to generating collective virtue. Unlike in the addition model, it is not required that group members are individually virtuous; in some cases, they might even be intellectually vicious.

3.1 *Mere Interaction*

The first and simplest version is when two or more individuals who work together stimulate or challenge each other – intentionally or not – to do better than they would have on their own. Interaction between individuals who lack virtue can consistently produce epistemically excellent outcomes.

You need not be a particularly virtuous individual for collaboration to awaken a competitive mindset, a desire to show your best self, an urge to impress other people, or at least to not let them down. If you have ever successfully coauthored a paper or cotaught a class, you should be able to recognize this phenomenon. There is a wealth of empirical research in cognitive and social psychology supporting the general idea that interaction and collaboration in a group lead people to modify their behavior in various ways (cf. Kelly et al. 2013).

Of course, not any combination of people who work together will automatically do so in virtuous ways; we are all familiar with stories about group processes gone terribly wrong. The claim is far more modest: sometimes, with the right combination of people and the right collaborative tasks and settings, people working together will excel even when they wouldn't have done so individually. Consider some schematic examples of actual intellectual virtues. By pooling ideas, asking critical and constructive questions, and building on each others' ideas, a group can become creative. When group members push each other to become clearer and more explicit and to think through potential criticisms, the group as a whole becomes intellectually careful and rigorous. If group members cheer each other on or refuse to give up first not to lose face, the group might persevere on a difficult task where individuals wouldn't have done so.

3.2 *Collective Virtue Out of Individual Vices*

The right combinations of individually intellectually *vicious* people could also form groups that possess intellectual virtue. Based on a range of empirical literature from biology, psychology, and organization science, Paul Smart (2018a, 2018b) defends this possibility by exploring what he calls 'Mandevillian intelligence':

Cognitive and epistemic properties that are typically seen as shortcomings, limitations or biases at the individual level can, on occasion, play a positive functional role in supporting the emergence of intelligent behavior at the collective level.

(Smart 2018b, 4171)

Many group intellectual tasks can be construed as a collective search through a space of doxastic possibilities: solving a problem, forming a hypothesis, making a decision, or forming a belief that's in accordance with the available evidence. As Smart points out, performing a collective search well requires striking a balance between exploration and exploitation. Unless the space of possible solutions is a simple ordered one, a successful search must explore the solution space far and wide in order to identify optimal solutions. The collective needs to look as

broadly as possible before it exploits group members' judgments to home in on a preferred solution. Somewhat surprisingly, fast and smooth information sharing among group members harms this process, because it leads to premature convergence on suboptimal solutions. A better balance between exploration and exploitation is achieved, Smart explains, when group members trust each other less, are individually dogmatic, or manifest cognitive biases and heuristics like confirmation bias, belief perseverance, the availability heuristic, etc. – in other words, when they exhibit individually vicious behaviors.

Smart's proposal dovetails with other strands of research. Modeling work in the philosophy of science has shown that one way for scientific communities to do a better job of converging on the truth under certain conditions is for individual scientists to start out with more extreme beliefs (Zollman 2010).³ Another way, again under certain conditions, is for scientists in a broader community to actively avoid approaches already taken by others (Weisberg and Muldoon 2009). Although this doesn't entail that individual scientists must be intellectually vicious for the community to be successful, it is clear that vices such as self-righteousness, narrow-mindedness, or arrogance might lead to extreme beliefs or might stimulate researchers to actively avoid approaches taken by others. Drawing on various strands of research in cognitive and evolutionary psychology, Hugo Mercier and Dan Sperber (2017) argue that human reason evolved for social use. Reasoning is meant to convince others, to justify our thoughts and actions to others, and to scrutinize others' justifications. Biases and limitations that may seem intellectually bad on the individual level produce epistemically successful interaction at the collective level by evolutionary design.⁴

3.3 *Structure and Culture*

On the third version of the interaction model, it isn't the mere interaction between individuals as such that leads to collective virtue, but the members' interaction with the group's formal or informal structure, rules, or culture.⁵ When a group of people work together in pursuit of some common goal, some forms of organization arise naturally: tasks are divided, people take on different roles, mutual expectations form, communication patterns develop spontaneously or are explicitly agreed upon, a system of sanctions might be put in place, and something less tangible like a group 'culture' or 'ethos' emerges. This is all the more true for established groups that work together over longer periods of time in a formal institutional setting such as an organization.

According to Seumas Miller (2010, 2019), organizations are systems of interdependent roles determined by four characteristic elements: structure, function, culture, and a system of sanctions. The same goes for subgroups within organizations, such as departments, teams, or other

collaborative groups. The group's *structure* consists of the differentiated roles in the group, typically defined by tasks or responsibilities for the person occupying the role, rules governing the performance of those tasks, and relations to the other roles. Established organizations typically have a formal structure that is explicitly specified, but a group has structure even in the absence of any explicit specification. In addition, there is the organization's informal structure, which may or may not diverge from its formally specified structure. Sometimes, people take on tasks that aren't officially part of their role or they follow unofficial rules in carrying out their tasks. An organization's *function* is what it is for; its official purpose. This is what the structure with its roles, tasks, and rules is supposed to accomplish – very generally put, to produce goods or render services of various kinds. An organization's or group's *culture* is its 'spirit' or 'ethos': the set of informal attitudes, values, norms, beliefs, desires, expectations, communication patterns, practices, etc. that pervade the group and that, together with its structure, determine its behavior and performance. Ideally, a group's formal structure and informal structure and culture are harmoniously aligned, but of course, this isn't always so. The final element is a *system of sanctions*, which captures what happens when group members violate the group's rules, norms, or values; anything from formal punishment to friendly corrections.⁶

A group's structure and culture (including its system of sanctions) can generate virtuous intellectual performance, regardless of the virtues or vices of individual group members. The *formal structure* of a group and the operative rules and responsibilities can encode intellectually virtuous practices by stimulating or prescribing actions and procedures that constitute virtuous behavior and by making nonvirtuous behavior more difficult. This can happen in any number of ways: from simple conventions and agreed-upon standard practices to a complete institutional system for dividing intellectual labor between different roles or sophisticated knowledge management systems. For instance, simple things like always letting a colleague proofread letters or memos or double-checking calculations before approving payments can reduce errors. This may not quite amount to intellectual virtue yet, but at least vices of carelessness and sloppiness are avoided. Senior management roles are often designed so that they complement each other and prevent one-sidedness: the tasks of a CEO require courage and steadfastness, whereas a Chief Risk Officer is supposed to be careful and temperate (de Bruin 2017, 117). A management team can become virtuous when it fills these roles with the right people. Or take the practice of preregistration in science (Nosek et al. 2018) and depositing data and analyses in the Open Science Framework (Foster and Deardorff 2017).⁷ By registering the design, methods, and hypotheses of a study before carrying it out and committing to sharing data openly, various kinds of questionable research practices are prevented, such as hypothesizing after the results are known (Kerr 1998),

p-hacking, or letting results disappear. When a research team commits to working by the principles of open science (and lives by those commitments), its research practices will become more careful and more reliable – more intellectually virtuous. In all of these examples, the responsibilities and tasks that belong to various roles in a team are specified so that the individuals fulfilling these roles will show behavior that is conducive to the epistemic excellence of the group, regardless of whether they would be individually so inclined. In other words, the group's structure produces collective virtue.

Group structure can be scaffolded by training, standard operating procedures, protocols, and various sorts of technological support. Pilots, for example, are required to use preflight checklists before taking off to make sure everything is safe (Degani and Wiener 1993). Checklists are widely used in other high-risk environments, too, where safety is of the highest concern. They eliminate unreliability that might otherwise ensue from human lapses of attention or forgetfulness. The use of redundancy and double-checking is another familiar procedure for spotting mistakes and thus promoting reliability. A fascinating historical example is the Mathematical Tables Project, which was devoted to tabulating higher mathematical functions before there were electronic computers (Grier 2013, Ch. 13). The project ran from 1938 to 1948 under the leadership of the Polish-American mathematician Gertrude Blanche. Mathematicians broke down the calculations for the values of complex functions into basic arithmetic operations, which were then carried out by as many as 450 unemployed individuals, to be subsequently aggregated into comprehensive tables. In order to secure impeccable reliability, which was crucial for the project's reputation, Blanche and her fellow mathematicians went to great lengths to weed out error: they employed six to eight different procedures to check each calculation (Grier 2013, 215)!

Many organizations use knowledge management tools such as document repositories, data warehouses, management information dashboards, intranets, etc.⁸ When implemented well, such tools ensure that the right information is easily accessible to the right people at the right time, so that the organization operates on the basis of reliable information and according to current procedures and practices. Easily accessible, reliable, and current information also enables groups to be transparent and to justify their actions when called upon to do so. This is conducive to or constitutive of intellectual virtues like honesty, responsibility, accountability, and truthfulness. An example from science is the massive open database with biochemical data at the European Bioinformatics Institute in Cambridge (Cook et al. 2018). Through the use of open (big) data, research teams in the life sciences can speed up discovery and enhance the reproducibility of their work.

A group's *culture* can contribute to its epistemic flourishing, too. Informal and implicit ideals, values, norms, practices, attitudes, beliefs,

communication patterns, and other attitudes and behaviors influence the group members in a multitude of ways. A nonexhaustive list of ways in which a group's culture might be embodied and expressed includes: are questions welcomed; are junior team members mentored; is there organizational support for learning and development; are there opportunities for creativity and out-of-the-box thinking; do team members (especially those in hierarchical relations) welcome feedback and criticism; do team members give each other credit; do group members take pride in being part of the group; do people experience the organization's overall goals as worth caring about; who are the group's role models; are work hours and compensation in proportion to the tasks and results that are expected; are successes celebrated; etc. All of these things set the tone and shape the group's ethos. They can create a group that has intellectual virtues, even when the individuals in the group are not particularly virtuous apart from the group.⁹

To give a concrete example, Richard Dawkins tells a charming anecdote that illustrates the idea of informal communal norms well. A senior scientist in the Oxford zoology department, where Dawkins was an undergraduate, had for years

passionately believed, and taught, that the Golgi Apparatus (a microscopic feature of the interior of cells) was not real: an artefact, an illusion. Every Monday afternoon it was the custom for the whole department to listen to a research talk by a visiting lecturer. One Monday, the visitor was an American cell biologist who presented completely convincing evidence that the Golgi Apparatus was real. At the end of the lecture, the old man strode to the front of the hall, shook the American by the hand and said – with passion – “My dear fellow, I wish to thank you. I have been wrong these fifteen years.” We clapped our hands red. ... The memory of the incident I have described still brings a lump to my throat.

(Dawkins 2006, 321)

It's easily relatable how events such as these can have a formative influence on a group: when a group member sets an example by an impressive display of virtuous behavior, others will want to live up to that ideal and strive to improve their own behavior in the image of that ideal. Especially when stories about an exemplar are often repeated or when little ritual-like practices are formed around it, they have a lasting influence. While such exceptional events shape group culture, day-to-day practices and dealings are arguably even more important. For another example from the domain of science, recent systematic research on research integrity is beginning to single out 'research climate' – which is basically synonymous with 'culture' in this context – as a key driver of responsible conduct of research and prevention of questionable research practices (Crain et al. 2013).

A case from the literature also illustrates the influence of culture on collective virtue (or lack thereof). Both Reza Lahroodi (2007) and Miranda Fricker (2010) discuss the example of a church committee that operates in a closed-minded fashion even when all of its members are individually open-minded. While this is an example of a group vice, the example could easily be reversed to be about a virtuous committee consisting of members with individual vices. In Lahroodi's words:

We can conceive of a church committee that is narrow-minded about gay rights as a group, while all or most of its members are open-minded about gay rights. As individuals, all or most members of the committee routinely resist their initial tendency to dismiss ideas favoring gay rights that are contrary to their own and to grant them enough plausibility to take them seriously. The group, however, moves in the opposite direction. It fails to assign any plausibility to a wide range of contrary views about gay rights, summarily dismisses them and does not consider them worthy of discussion, let alone adoption.

(Lahroodi 2007, 287)

What explains the committee's behavior, as Lahroodi describes the example, is the interplay between two sorts of factors: first, 'commitment to certain standards, including standards for satisfactory discharge of the group's tasks, standards for good evidence or good reasoning about subjects relevant to the group's tasks, and so on'; and second, 'the pressure on members to reinforce their group membership by performing conforming behavior... [Group members] may want others to think they are towing the church line on this issue' (ibid., 288). Both of these factors form part of the group culture.

Miranda Fricker (2010) uses Christine Korsgaard's (1996) notion of practical identities to make sense of such a group dynamic. Following Korsgaard, she notes how people have various sorts of identities, corresponding with the different roles they occupy in their personal, social, and professional life: depending upon the circumstances and occasion, one can engage a situation as a parent, as a citizen, as a party member, as an employee, as a team member, etc. Some of these practical identities arise from group membership and the values and norms associated with the identity are set by the group ethos. Practical identities can express themselves in different beliefs, acceptances, utterances, and actions, which can in turn influence what other group members do in their roles as group members, thus creating a distinctly collective dynamic. A jury member in a legal court might, because of personal prejudice or hasty judgment, be individually convinced that the accused is guilty, but nonetheless realize that, *qua* jury member, she ought to refrain from judgment and wait until all the evidence has been presented and deliberations

are under way. In so far as all jury members wear their practical identities in this way, the jury as a group can be fair-minded and intellectually responsible.

In a later paper, Fricker (2020) employs Margaret Gilbert's (2014) notion of joint commitment to analyze group ethos.¹⁰ While this is a fruitful idea, analyses of collective virtue are not wedded to the joint commitment model.¹¹ For example, writing about social knowledge, Alexander Bird draws on the Durkheimian notion of organic solidarity to characterize the way in which some groups are bound together. Organic solidarity, he writes, 'involves bonds that arise out of difference, primarily the interdependence brought about by the division of labor. The key feature of the division of labor is that individuals and organizations depend on others who have different skills and capacities' (Bird 2010, 37). Bird is explicit that groups bound by organic solidarity need not take on any joint commitments in Gilbert's sense. Even so, such groups have an ethos, too, which makes them function well, neutrally, or badly from an epistemic point of view. Joint commitment may be a fine conceptual tool for understanding what group culture or ethos can be, but we don't need to limit our theoretical options here.¹²

To sum up the ideas from this section: the interactive model of collective virtue has three versions. First, the mere interaction between collaborating people who are individually lacking in individual virtue can produce epistemic excellence at the group level. Second, the right combinations of individual vices can produce epistemic excellence through interaction. And third, group structure and culture can nudge, coax, push, or require individuals to behave and interact in ways that make the group as a whole flourish epistemically, regardless of the epistemic qualities of the group members.

4 Emergence

To introduce the third model, I need to draw attention to an implicit assumption in the discussion so far. It is that there is a single set of intellectual virtues, which can be had by individuals and groups alike. The examples so far included familiar ones from the virtue epistemology literature: reliability, love of knowledge, responsiveness to evidence, open-mindedness, perseverance, teachability, creativity, courage, etc. The third model – admittedly the most speculative of the three – turns on the insight that when we relax the assumption that all virtues can be possessed by both groups and individuals, there is theoretical space for intellectual virtues that *only* collectives can possess.¹³ Perhaps groups can possess intellectual virtues which no individual could possess: *exclusively collective virtues*. This suggestion does not require positing any mysterious mechanisms or group-level mentality or agency. The mechanisms through which exclusively collective virtues could emerge are similar to

those in the interaction model: interaction between people who individually lack virtue, well-ordered interaction between individual vices, or interaction between group structure and culture, and individual character and behavior. The difference with the interaction model lies in the kind and nature of the virtue itself, not the way in which it is produced.

To warm up to the idea that there can be exclusively collective virtues, let's start with two related phenomena which are better documented. First, it is uncontroversial that, under the right circumstances, collectives can outperform individuals. This is also true in the epistemic realm. Teams can work faster and more reliably than individuals; they can be better at generating new ideas; they can persevere longer (e.g., by dividing up labor); they can bring a greater number of diverse perspectives to an issue; etc. A specific example is the 'diversity trumps ability' theorem (Hong and Page 2004), which shows that teams consisting of sufficiently diverse problem solvers can outperform individual experts and even teams of experts. While individuals can surely bring a number of different perspectives to a problem, a team of diverse individuals can do so to a much higher degree and this theorem shows that, at least for some tasks, diversity matters more than expertise. So, for at least some of the virtues that individuals and collectives can both have, it is possible for collectives to have those virtues to a significantly greater degree than any individual could.

This observation makes a weak version of the third model plausible: for some virtues that both individuals and collectives can have, collectives can have them to a greater degree than any individual could. At least, then, there are exclusively collective virtues in the sense that there are degrees of virtue possession exclusive to collectives – *quantitatively exclusively collective virtues*, we might call them.

A second phenomenon suggests that there is room for a stronger version of the third model. There may be kinds of virtues that are unique to collectives – *qualitatively exclusively collective virtues*. To support this, consider the concept of superdiversity. Introduced by the sociologist Steven Vertovec (2007), this concept characterizes geographical regions or cities that have high numbers of different immigrant groups or people of different ethnicities and, as a result, lack any homogenous majority groups. It has been claimed that superdiversity is conducive to innovation and economic growth (Ozgen et al. 2012) and that it can reduce intergroup tensions and prejudice (see Foner et al. 2019 for discussion and references).¹⁴ While superdiversity and its effects are not intellectual virtues and cities and communities are not the sort of groups that form the focus of this chapter, it is clear that superdiversity is, by definition, a feature that only collectives can possess. It is structurally similar to the kinds of exclusively collective virtues I am trying to delimit here.

Of course, the question is whether there are compelling examples from the epistemic realm that fit the bill of this third model: characteristics of

groups that are conducive to or constitutive of epistemic flourishing and that only groups can have. A conservative approach to this question is to scrutinize detailed analyses of familiar individual virtues and to ask whether there are perhaps *forms* of these virtues that only groups can possess. In so far as the genus-species distinction applies to intellectual virtues, this is a promising avenue. Individuals and groups can both have virtues like intellectual humility, open-mindedness, creativity, perseverance, etc., but the specific form they take in individuals and collectives might differ. Virtues bifurcate into an individual and a collective form.

Consider the virtue of intellectual autonomy or self-governance. Individuals can be intellectually autonomous by thinking for themselves and deciding for themselves whom to trust. But any individual has only her own mind and cognitive resources to accomplish this. This is different for groups. First, because, unlike individuals, groups aren't 'all-purpose cognizers.' Groups only think and reason in so far as this is relevant to their function and purpose. Second, some groups have designated individuals to work on subtasks that are relevant to the overall intellectual task the group is engaged in. Hence, for groups, fairly radical forms of autonomy can be feasible and desirable: some groups can truly think fully for themselves and rely (almost) exclusively on their own resources, without trusting others outside the group. Along the same lines, Byerly and Byerly (2016) suggest that self-regulation can take on a distinctively collective form. Plausibly, self-regulation is an element of intellectual autonomy in so far as autonomy involves the group regulating the actions of its members. This form of self-regulation doesn't exist at the individual level, simply because there are no members whose behaviors can be regulated.

Something analogous can be said about cognitive diversity. An individual can have cognitive diversity by mastering various thinking styles, drawing on different experiences, and having different practical or social identities which she can bring to bear on questions. But clearly, a group can host a wider range of cognitive diversity by having members with radically different life histories, socio-economic, religious, or political backgrounds, and diverse lived experiences. Cognitive diversity might not be an intellectual virtue in and of itself, but it is certainly instrumental to virtues such as problem-solving capacity or creativity. Perhaps, then, there is a distinctively collective form of creativity.

Finally, a more radical approach looks for collective virtues that are truly unique in kind, that is, not just a species of the same genus as individual virtues, but such that individuals cannot have them. Byerly and Byerly (2016) propose that solidarity might be an example of such an exclusively collective virtue. This, however, isn't an intellectual virtue. A possible example from the intellectual realm involves the qualities of a group involved in fostering and cultivating mutual empathetic understanding. Michael Hannon (2020) argues that democratic deliberation

might be good for an empathetic understanding of other people. Understanding others, he writes, requires ‘that we be willing to listen to them. More than this, however, it requires the ability to “take up” another person’s perspective. We must be able to see the other person’s point of view’ (2020, 598). Hannon cites empirical evidence showing that, in the right circumstances, groups composed of diverse members who engage in deliberation indeed develop stronger empathetic understanding, which subsequently also increases outgroup empathy (Mutz 2006; Morell 2010; Grönlund et al. 2017). While mutual empathetic understanding might not be a strictly veritistic epistemic goal, it is nonetheless an epistemic goal, argues Hannon. It facilitates more accurate opinions about other people and is a precondition for rational deliberation, which may, in turn, enable better truth-tracking in political, moral, and religious matters.

Obviously, such mutual empathetic understanding is not a feature that individuals can possess. Only groups that are sufficiently diverse and that have a structure and culture that facilitate respectful dialogue will reap these epistemic benefits. This, then, is reason to think that the features that make groups good at cultivating empathetic understanding constitute a qualitatively exclusively collective intellectual virtue. Perhaps, then, we can call it the virtue of mutual empathetic understanding.

In conclusion, the third model for collective virtue presents the possibility that groups possess intellectual virtues that individuals cannot have. Either by having a familiar virtue to a greater degree than any individual could – a quantitatively exclusively collective virtue – or by having a virtue that only groups can have – a qualitatively exclusively collective virtue.

5 Conclusion

I want to close by offering two suggestions for future research on these three models for collective intellectual virtue, which can advance this new branch of collective epistemology and virtue theory. The first is to dive into the issues that I bracketed for the purposes of this chapter: (a) whether collective virtues really exist and (b) whether groups can have both reliabilist and responsibilist virtues. This requires connecting the three models I have outlined and discussed here to the extensive literature on intellectual virtue. To address (a), a general account is needed of when a virtue is a genuinely collective one. Such an account can then be compared to the three models and their different versions I have outlined above. For (b) we need developed accounts of both reliabilist (Sosa 2007; Greco 2010) and responsibilist (Montmarquet 1993; Zagzebski 1996; Baehr 2006; Roberts and Wood 2007) virtues, which can then be used to identify the conditions which groups must meet in order to possess both kinds of virtues. Particularly for responsibilist virtues, which are

often held to require virtuous motivation, this might require a further account of group motivation.

The second suggestion is to develop the three models in more empirical detail, by looking at research from social psychology, sociology, organization science, etc. on group dynamics and performance to identify the specific and measurable conditions under which groups flourish epistemically. The chapters in this volume by Ryan Byerly and Marco Meyer already take important steps in this direction. The models as described above are largely schematic and leave open questions like: what combinations of virtues work together well; which specific vices can produce which collective virtues; what are good organizational structures, cultures, and support systems to cultivate collective intellectual virtues; and so on. Even though the concept of collective intellectual virtue is not widely used in social science, a lot of extant research may well be highly relevant to answering these questions.

Notes

- 1 See Fletcher (2018) for the basics of the science and technology.
- 2 For these, see the extensive literature on the Condorcet Jury Theorem (Goodin and Spiekermann 2018), the ‘Miracle of Aggregation’ (Converse 1990; Page and Shapiro 1993), Scott Page’s ‘The Crowd Beats the Average Law’ (Page 2008, 209), and information / prediction markets (Wolfers and Zitzewitz 2004; Tetlock and Gardner 2015; Dana et al. 2019).
- 3 See Frey and Šešelja (2020), however, for robustness worries about Zollman’s results.
- 4 Sloman and Fernbach (2017) similarly argue for a collectivist account of cognition.
- 5 In practice, it will be nearly impossible to tease these ‘mere interaction’ apart from ‘culture and structure’: when two or more people collaborate over some period of time, a certain culture and structure inevitably emerge. Moreover, culture and structure aren’t separate from individual interaction. On the contrary, they manifest themselves through individual interactions over time. For analytical purposes, though, it is helpful to focus on structure and culture as separate entities with causal influence on a group’s behavior.
- 6 I’m inclined to think that a system of sanctions can be construed as an element of the group’s structure and culture, but I’m following Miller in listing it separately.
- 7 See also: <https://osf.io>.
- 8 Syed et al. (2018, Part III) provides a wide range of examples and discussion.
- 9 Needless to say, all of the above can conspire to produce vice, too. Stories about dysfunctional organizations, cultures of fear, workplace bullying, harassment, incompetent management, implicitly enforced inequality, silenced or smothered voices, etc. are unfortunately all too familiar.
- 10 In the already cited earlier paper, Fricker (2010) also used this notion to analyze group *motivation* in order to offer an account of virtuous collective motivation as part of a responsibilist account of collective virtue.
- 11 Byerly and Byerly (2016) offer further systematic reasons against using the theoretical apparatus of joint commitment to analyze collective virtue.
- 12 In fact, joint commitment might be more appropriate for analyzing group *structure*. Roles, tasks, responsibilities, and other elements of a group’s

structure are usually explicitly discussed and agreed upon by group members and, in institutionalized settings, they are often specified in official documents. This lends itself readily to an analysis in terms of joint commitments where group members express their willingness to commit to their respective roles and responsibilities in the group while knowing that others have also expressed such willingness.

13 Byerly and Byerly (2016, §3) also explore this suggestion.

14 Note that the theoretical usefulness of the concept is not uncontroversial (see, e.g., Deumert 2014; Pavlenko 2018).

References

- Baehr, Jason. 2006. "Character, Reliability and Virtue Epistemology." *The Philosophical Quarterly* 56 (223): 193–212. <https://doi.org/10.1111/j.1467-9213.2006.00437.x>.
- Battaly, Heather. 2008. "Virtue Epistemology." *Philosophy Compass* 3 (4): 639–663. <https://doi.org/10.1111/j.1747-9991.2008.00146.x>.
- Bird, Alexander. 2010. "Social Knowing: The Social Sense of 'Scientific Knowledge.'" *Philosophical Perspectives* 24 (1): 23–56. <https://doi.org/10.1111/j.1520-8583.2010.00184.x>.
- Bruin, Boudewijn de. 2017. *Ethics and the Global Financial Crisis*. Cambridge: Cambridge University Press.
- Byerly, T. Ryan, and Meghan Byerly. 2016. "Collective Virtue." *The Journal of Value Inquiry* 50 (1): 33–50. <https://doi.org/10.1007/s10790-015-9484-y>.
- Converse, Philip. 1990. "Popular Representation and the Distribution of Information." In *Information and Democratic Processes*, edited by John A. Ferejohn and James H. Kuklinski, 369–389. Chicago: University of Illinois Press.
- Cook, Charles E., Mary T. Bergman, Guy Cochrane, Rolf Apweiler, and Ewan Birney. 2018. "The European Bioinformatics Institute in 2017: Data Coordination and Integration." *Nucleic Acids Research* 46 (D1): D21–D29. <https://doi.org/10.1093/nar/gkx1154>.
- Crain, A. Lauren, Brian C. Martinson, and Carol R. Thrush. 2013. "Relationships between the Survey of Organizational Research Climate (SORC) and Self-Reported Research Practices." *Science and Engineering Ethics* 19 (3): 835–850. <https://doi.org/10.1007/s11948-012-9409-0>.
- Dana, Jason, Pavel Atanasov, Philip E. Tetlock, and Barbara Mellers. 2019. "Are Markets More Accurate than Polls? The Surprising Informational Value of 'Just Asking.'" *Judgment and Decision Making* 14 (2): 135–147.
- Dawkins, Richard. 2006. *The God Delusion*. Boston: Houghton Mifflin Harcourt.
- Degani, Asaf, and Earl L. Wiener. 1993. "Cockpit Checklists: Concepts, Design, and Use." *Human Factors* 35 (2): 345–359. <https://doi.org/10.1177/001872089303500209>.
- Deumert, Ana. 2014. "Digital Superdiversity: A Commentary." *Discourse, Context & Media* 4–5: 116–120. <https://doi.org/10.1016/j.dcm.2014.08.003>.
- Deylin, Hannah. 2019. "Black Hole Picture Captured for First Time in Space Breakthrough." *The Guardian*, April 10, 2019. <http://www.theguardian.com/science/2019/apr/10/black-hole-picture-captured-for-first-time-in-space-breakthrough>.

- Epstein, Brian. 2019. "What Are Social Groups? Their Metaphysics and How to Classify Them." *Synthese* 196 (12): 4899–4932. <https://doi.org/10.1007/s11229-017-1387-y>.
- Fletcher, Seth. 2018. "How Do You Take a Picture of a Black Hole? With a Telescope as Big as the Earth." *The New York Times*, October 4, 2018. <https://www.nytimes.com/2018/10/04/magazine/how-do-you-take-a-picture-of-a-black-hole-with-a-telescope-as-big-as-the-earth.html>.
- Foner, Nancy, Jan Willem Duyvendak, and Philip Kasinitz. 2019. "Introduction: Super-Diversity in Everyday Life." *Ethnic and Racial Studies* 42 (1): 1–16. <https://doi.org/10.1080/01419870.2017.1406969>.
- Foster, Erin D., and Ariel Dearnorff. 2017. "Open Science Framework (OSF)." *Journal of the Medical Library Association : JMLA* 105 (2): 203–206. <https://doi.org/10.5195/jmla.2017.88>.
- Frey, Daniel, and Dunja Šešelja. 2020. "Robustness and Idealizations in Agent-Based Models of Scientific Interaction." *The British Journal for the Philosophy of Science* 71 (4): 1411–1437. <https://doi.org/10.1093/bjps/axy039>.
- Fricke, Miranda. 2010. "Can There Be Institutional Virtue?" In *Oxford Studies in Epistemology, Volume 3*, edited by Tamar Szabo Gendler and John Hawthorne, 235–252. New York: Oxford University Press.
- . 2020. "Institutional Epistemic Vices: The Case of Inferential Inertia." In *Vice Epistemology*, edited by Ian James Kidd, Heather Battaly, and Quassim Cassam, 89–107. London: Routledge.
- Gilbert, Margaret. 1987. "Modelling Collective Belief." *Synthese* 73 (1): 185–204.
- . 2014. *Joint Commitment: How We Make the Social World*. New York: Oxford University Press.
- Gilbert, Margaret, and Daniel Pilchman. 2014. "Belief, Acceptance, and What Happens in Groups: Some Methodological Considerations." In *Essays in Collective Epistemology*, edited by Jennifer Lackey, 189–212. New York: Oxford University Press.
- Goodin, Robert E., and Kai Spiekermann. 2018. *An Epistemic Theory of Democracy*. New York: Oxford University Press.
- Greco, John. 2010. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press.
- Grier, David Alan. 2013. *When Computers Were Human*. Princeton, NJ: Princeton University Press.
- Grönlund, Kimmo, Kaisa Herne, and Maija Setälä. 2017. "Empathy in a Citizen Deliberation Experiment." *Scandinavian Political Studies* 40 (4): 457–480. <https://doi.org/10.1111/1467-9477.12103>.
- Hannon, Michael. 2020. "Empathetic Understanding and Deliberative Democracy." *Philosophy and Phenomenological Research* 101 (3): 591–611. <https://doi.org/10.1111/phpr.12624>.
- Hong, Lu, and Scott E. Page. 2004. "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers." *Proceedings of the National Academy of Sciences* 101 (46): 16385–16389. <https://doi.org/10.1073/pnas.0403723101>.
- Kelly, Janice R., Megan K. McCarty, and Nicole E. Iannone. 2013. "Interaction in Small Groups." In *Handbook of Social Psychology*, edited by John

- DeLamater and Amanda Ward, 413–438. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6772-0_14.
- Kerr, Norbert L. 1998. “HARKing: Hypothesizing After the Results Are Known.” *Personality and Social Psychology Review* 2 (3): 196–217. https://doi.org/10.1207/s15327957pspr0203_4.
- Korsgaard, Christine M. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Lackey, Jennifer, ed. 2014. *Essays in Collective Epistemology*. Oxford and New York: Oxford University Press.
- . 2021. *The Epistemology of Groups*. New York: Oxford University Press.
- Lahroodi, Reza. 2007. “Collective Epistemic Virtues.” *Social Epistemology* 21 (3): 281–297. <https://doi.org/10.1080/02691720701674122>.
- . 2019. “Virtue Epistemology and Collective Epistemology.” In *The Routledge Handbook of Virtue Epistemology*, edited by Heather Battaly, 407–419. London: Routledge.
- Mercier, Hugo, and Dan Sperber. 2017. *The Enigma of Reason*. Cambridge, MA: Harvard University Press.
- Miller, Seumas. 2010. *The Moral Foundations of Social Institutions: A Philosophical Study*. Cambridge: Cambridge University Press.
- . 2019. “Social Institutions.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2019. <https://plato.stanford.edu/archives/sum2019/entries/social-institutions/>.
- Montmarquet, James A. 1993. *Epistemic Virtue and Doxastic Responsibility*. Lanham, MD: Rowman & Littlefield.
- Morell, Michael E. 2010. *Empathy and Democracy: Feeling, Thinking, and Deliberation*. University Park, PA: Penn State University Press.
- Mutz, Diana C. 2006. *Hearing the Other Side: Deliberative versus Participatory Democracy*. Cambridge: Cambridge University Press.
- Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. “The Preregistration Revolution.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (11): 2600–2606. <https://doi.org/10.1073/pnas.1708274114>.
- Ozgen, Ceren, Peter Nijkamp, and Jacques Poot. 2012. “Immigration and Innovation in European Regions.” In *Migration Impact Assessment*, edited by Peter Nijkamp, Jacques Poot, and Mediha Sahin, 261–298. Cheltenham: Edward Elgar.
- Page, Scott E. 2008. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton, NJ: Princeton University Press.
- Page, Benjamin I., and Robert Y. Shapiro. 1993. “The Rational Public and Democracy.” In *Reconsidering the Democratic Public*, edited by George E. Marcus and Russell L. Hanson, 35–64. University Park, PA: Pennsylvania State University Press.
- Pavlenko, Aneta. 2018. “Superdiversity and Why It Isn’t: Reflections on Terminological Innovation and Academic Branding.” In *Sloganization in Language Education Discourse*, edited by Barbara Schmenk, Stephan Breidbach, and Lutz Küster, 142–168. Bristol: Multilingual Matters. <https://doi.org/10.21832/9781788921879-009>.

- Quinton, Anthony. 1975. "Social Objects." *Proceedings of the Aristotelian Society* 76: 1–27.
- Ridder, Jeroen de. 2014. "Epistemic Dependence and Collective Scientific Knowledge." *Synthese* 191: 37–53.
- Ritchie, Katherine. 2015. "The Metaphysics of Social Groups." *Philosophy Compass* 10 (5): 310–321. <https://doi.org/10.1111/phc3.12213>.
- Roberts, Robert C., and W. Jay Wood. 2007. *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford: Clarendon Press.
- Slooman, Steven, and Philip Fernbach. 2017. *The Knowledge Illusion: Why We Never Think Alone*. New York: Riverhead.
- Smart, Paul R. 2018a. "Mandevillian Intelligence: From Individual Vice to Collective Virtue." In *Socially-Extended Epistemology*, edited by Joseph Adam Carter, Andy Clark, Jesper Kallestrup, Spyridon Orestis Palermos, and Duncan Pritchard, 253–274. Oxford: Oxford University Press.
- . 2018b. "Mandevillian Intelligence." *Synthese* 195 (9): 4169–4200. <https://doi.org/10.1007/s11229-017-1414-z>.
- Sosa, Ernest. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford: Clarendon Press.
- Syed, Jawad, Peter Murray, Donald Hislop, and Yusra Mouzoughi, eds. 2018. *The Palgrave Handbook of Knowledge Management*. Cham: Palgrave Macmillan. <https://www.palgrave.com/gp/book/9783319714332>.
- Tetlock, Philip E., and Dan Gardner. 2015. *Superforecasting: The Art and Science of Prediction*. New York: Crown.
- Tuomela, Raimo. 1992. "Group Beliefs." *Synthese* 91 (3): 285–318. <https://doi.org/10.1007/BF00413570>.
- . 2004. "Group Knowledge Analyzed." *Episteme* 1 (2): 109–127. <https://doi.org/10.3366/epi.2004.1.2.109>.
- Vertovec, Steven. 2007. "Super-Diversity and Its Implications." *Ethnic and Racial Studies* 30 (6): 1024–1054. <https://doi.org/10.1080/01419870701599465>.
- Weisberg, Michael, and Ryan Muldoon. 2009. "Epistemic Landscapes and the Division of Cognitive Labor." *Philosophy of Science* 76 (2): 225–252. <https://doi.org/10.1086/644786>.
- Wolfers, Justin, and Eric Zitzewitz. 2004. "Prediction Markets." *Journal of Economic Perspectives* 18 (2): 107–126. <https://doi.org/10.1257/0895330041371321>.
- Zagzebski, Linda T. 1996. *Virtues of the Mind*. Cambridge: Cambridge University Press.
- Zollman, Kevin J. S. 2010. "The Epistemic Benefit of Transient Diversity." *Erkenntnis* 72 (1): 17–35. <https://doi.org/10.1007/s10670-009-9194-6>.

12b Commentary from Kate Devitt

Bounded Epistemology: Normative Versus Descriptive Considerations in Three Models of Collective Virtues

This chapter begins with two evocative examples, on the one hand, the collective virtue of the creation of the image of a black hole in 2019 and on the other hand, the collective vice of a government falsely accusing parents of defrauding the childcare benefit system. I want to interrogate these examples to illuminate a normative concept of collective epistemology based on bounded human faculties. What's telling about both these examples is the use of technology to enable human epistemic aims whether it is the use of algorithms to integrate multiple data sources or government dependence on automated decision making. Technologies have always enabled humans to go beyond limited faculties and to mitigate their biases, whether this is through physical path-creation, the physical arrangement of tools (Sterelny 2012), or an international network of radio telescopes (Akiyama et al. 2019). I believe acknowledging human-technological systems is key to unlocking normative versus descriptive models of collective epistemic virtues.

Collective human epistemic endeavor comprises of both 'making the best' of our inherent limits as well as striving for and building better knowledge-generating tools and systems. Thus, our theory of collective virtues ought to be grounded in a reasonable theory of virtuousness, one that only asks of humans what they might possibly achieve and out-sources the rest to artifacts and methods.

I'm influenced by the work on bounded rationality (Gigerenzer 1991; Gigerenzer & Goldstein 1996). In bounded rationality, humans are rational – both justified and accurate – in their decision-making when they are within appropriate contexts and information environments, that is, ones they have evolved to think about. Gigerenzer argues that humans can be very good at reasoning if information is presented, say using frequencies rather than probabilities. Perhaps this is because we evolved in contexts where mathematical information was available in visible quantities (e.g., herds of bison roaming), rather than abstracted to concepts such as percentages or likelihoods. I don't want to get into too

much detail on their theory, but I think the main principles (contextual virtue) might apply when considering models for epistemic virtues.

Consider the summative model, where a group does well epistemically when all or most of its members are epistemically excellent. Addition is clearly a fantastic model where agents are facsimiles of each other, and able to be reliable and well-motivated, say a fleet of self-driving cars. As a collective, Tesla cars gather data about their environments as they drive and data about their drivers' behaviors. Cars provide data for the collective good of the fleet. Each car benefits from the addition of the data and insights from every other car. There is a collective virtue beyond the virtue of each car in the centralized algorithms that integrate the data and create models of the world. Teslas suffer none of the problems of open-mindedness leading to intellectual feebleness. Teslas can sense all data of relevance and remain vigilant as this data is processed. So, perhaps a summative model is good for artificial agents, where epistemic models can be programmed into systems that hold epistemic behaviors consistent with normative goals. But, the limits of each human and their biases suggest we need a different model to understand virtue among humans as well as telegraph what humans ought to do and be with technologies.

Humans are diverse and therefore will have different capabilities to add to the collective. A group might have virtues that work fine on an individual level, but as a collective, amplify group think or prematurely reduce innovation. The interaction model supposes that collective virtue can be produced by the interaction of individuals without these individuals necessarily possessing virtue themselves. If one adopts the motivation model of virtue, then this is understandable. I'm not sure if this theory bears out in the competence model of virtue. For, surely epistemically incompetent teams, even with good motivation, will not drive good collective epistemic outcomes? The models of trust emerging in the business management literature may go some way in articulating why both integrity (motivation, character, honesty) and competence (experience, skills, reliability) are important in building virtuous trustworthy teams (see Connelly et al. 2015).

In my proposed bounded epistemic model, individuals will still need to have some virtue in some contexts to contribute to group virtue. Diverse human teams will have different virtues useful in different contexts. Human teams with the right processes, methods and culture will harness these competences when they are needed and deprioritize individuals when they are 'out of their depth' epistemically. For example, open-minded people should be prioritized in creative ideation tasks, systems-thinkers should be prioritized in analysis and fastidious people prioritized in task completion. As de Ridder points out, sometimes the vices of individuals might debias the collective – vices being virtuous in some cases. Group culture and processes may also make up for individual

vice and contexts that are likely to yield vice. Still, if two groups were compared and the individuals of one group had more virtue than the individuals of another group and both groups had group-wide systems to overcome limits; the group with individuals with greater competence and integrity would do better epistemically than the group that did not.

Given that humans have always used technology and tools to augment knowledge-seeking, it makes sense that the right model of collective virtue ought to be grounded in our best theories of normative epistemology and could include emergent virtues that do not reside in individual humans or artificial agents alone. Once we are agreed on normative epistemic models, then we can seek to implement them, accepting the bounded epistemic competence and integrity of humans.

References

- Akiyama, K., Alberdi, A., Alef, W., Asada, K., Azulay, R., Baczko, A. K., ... Ramakrishnan, V. (2019). First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole. *The Astrophysical Journal Letters*, 875(1), L4.
- Connelly, B. L., Crook, T. R., Combs, J. G., Ketchen, D. J., & Aguinis, H. (2015). Competence- and Integrity-Based Trust in Interorganizational Relationships: Which Matters More? *Journal of Management*, 44(3), 919–945. doi:10.1177/0149206315596813
- Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond “Heuristics and Biases”. *European Review of Social Psychology*, 2, 83–115. doi:10.1080/14792779143000033
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review*, 103(4), 650–669. doi:10.1037/0033-295X.103.4.650
- Sterelny, K. (2012). *The Evolved Apprentice*. MIT Press.

12c Commentary from Heidi Grasswick

Jeroen de Ridder's 'Three models for "Collective Virtues"' offers a helpful map of a variety of ways one can understand the possibility of epistemic collective virtues. His claims are richly supported by his inclusion of not only philosophical work on collective virtues, but also interesting empirical work offering examples of how certain collective practices and social structures can support specific epistemic pursuits undertaken in collaborative group contexts.

Importantly, De Ridder sets out to side-step several sticking points in the debates concerning the possibility of collective virtues, including whether collective epistemic states simply amount to shorthand for individual epistemic states, what kinds of groups (if any) can actually possess virtues, and what the ultimate nature of virtue amounts to. I appreciate de Ridder's focus on the constructive task of 'understanding the different things we might mean when we use the language of intellectual virtue to evaluate group epistemic performance' (p. 368), as a way of moving forward in coming to understand the role of groups and communities (and their social structures) in epistemic practices. If in fact in different contexts we mean different things when we appeal to collective intellectual virtues, it is likely the case that each model de Ridder offers tells us something different about how collectives can play a role in epistemic practices and performance, with there being no need to settle on one particular model.

De Ridder sets out to articulate three different models of collective virtue, all of which seem to have a place in understanding the ways in which groups may take part in the virtues of knowing (and the vices). I draw attention to just a handful of points in considering the implications of these models and where we go from here in the study of collective epistemic virtues.

The first model discussed is what de Ridder calls the additive model, which expresses the idea that a group 'does well epistemically when all or most of its members are epistemically excellent' (369). In his gloss, de Ridder notes that virtuous individuals who collaborate can exhibit

virtuous behavior as a team (369), having stipulated already that the focus of his analysis will be on ‘collaborative communities’ where people working together to achieve a goal.

There is of course a tendency for those interested in theorizing collective virtues to downplay the importance of this particular model, though it will be an attractive position for those who wish to emphasize the need to ultimately understand virtues at an individual level, even while recognizing that we often produce knowledge in groups. But even within de Ridder’s description, we can identify both a weak and a strong sense of the additive or summative account. This is because he seems to identify two different ways in which we might refer to a collective virtue that is animated in a high proportion of the collective’s members. In one sense, we may claim a collective virtue simply in virtue of the fact that its members (or most of them) exhibit the virtue. This might be most weakly understood as a metaphorical understanding of a collective virtue. We talk as though the group ‘has’ the virtue of being intellectually curious because when we look at its members they (or at least most of them) are performing their work with intellectual curiosity. Yet de Ridder also expresses the additive model in a slightly different way when he suggests that ‘Scholars who are individually intellectually perseverant can stimulate each other to become even more perseverant when they collaborate. An open-minded thinker and an intellectually generous one might make a great teaching team, and so on’ (369). This situation suggests a slightly stronger model than the initial description, and it is less easy to dismiss as just metaphorical talk about collective virtue. It is still an additive account given that it is the virtuous performance of most of the individuals that is leading to the claim that the collective is virtuous in the said ways, yet it is because of the collaboration and the interaction between the individuals that each individual performs even more virtuously than they would otherwise. In essence, this latter situation has started to build in the important role of interaction that is the central driver of de Ridder’s second model (interaction), though it remains additive given the limited ways in which it does this.

Though this may seem to be a small distinction I have made here, it is noteworthy, given that additive accounts are sometimes pre-emptively dismissed by those working on collective virtues. Yet there is something to be said for attending to virtuous (not to mention vicious) feedback loops between individuals and the collectives within which they are working. Though a primary interest may be to get a handle on what we mean when we claim collective virtues, the ways in which they can help support individual virtuous behavior should not be lost, particularly when we admit that there can often be situations of positive feedback loops.

The second model de Ridder discusses – the interactive model – emphasizes several different kinds of interaction from which we might

understand a collective virtue to emerge without the individual members possessing that virtue themselves (or not in the same degree). The group's performance can be improved by members stimulating and challenging each other, but interestingly this model also captures interaction with the group's structures and culture. Recognizing these as features of our communities and institutions rather than viewing them as just the context within which individual-to-individual interactions occur is an important move towards grasping the significance of collectivities in epistemic performance.

More interesting still, in his discussion of the interactive model, de Ridder points to ways in which collective virtue could emerge out of individual vices. This feature brings with it the potential repercussion that in certain circumstances we might actually *want* individuals to have particular intellectual vices that can be balanced out within a working team. If this is the case, we might need to reassess how we think about individual virtues and vices if we accept that large amounts of epistemic work are in fact done in collaborative, community or team settings, with the ultimate goal being high-quality epistemic work coming out of the team, with little importance given to the individuals' performances. De Ridder draws on several interesting studies here to motivate the claim that individual vices might result in a virtuous collective, and at the end of his chapter, he calls for the further development of these models to include more of such empirical work. This is of course a very large task, but it does seem an important one. Ultimately we'll need to know under what circumstances we'd want certain types of virtues displayed in the collaborating individuals, and under what circumstances this might be less important or even antithetical to the ultimate goal of highly successful collective epistemic work.

De Ridder's third model presents the possibility of there being some emergent collective virtues that have no corresponding individual version. He notes that this is the most speculative model, but it is also the most provocative, asking us to look hard at what features of a collective contribute to epistemic success in different contexts rather than simply expanding on accounts of relatively well understood individual virtues. The very fact that his account of potential emergent collective virtues proceeds cautiously, first making the case for the possibility of 'quantitatively exclusive collective virtues', followed by exploring 'qualitatively exclusive' yet nonepistemic collective virtues, and finally thinking about specific forms of epistemic virtues that might properly apply to collectives suggests that although this model may be currently underdeveloped, it could have exciting potential to disrupt the intellectual trajectory we typically see within epistemology: a trajectory in which discussions of collective knowing are added onto a pre-existing individualized scaffolding rather than radically rethinking (as least some of) what we need in order to collectively know well.

In sum, the three models de Ridder sets out are interesting of themselves, but they also suggest new ways of coming at those core questions concerning the nature of and requirements of collective virtues that he sidesteps at the outside. For example, I interpret his models as potentially inclusive of each other: they could be, and likely are, embodied by some of the same collaborative communities. A community might exhibit a certain collective epistemic virtue by way of both the additive, and interactive models. However, not all aspects of these models are necessarily compatible. If in certain contexts, or if in relation to certain goals, we need individuals with various vices in order to obtain an optimal collectively virtuous performance, that will not be compatible with the additive model (at least with respect to the particular virtue in question), and it may not be compatible with certain other ways in which the interactive model articulates the effects of interaction in creating ‘more virtue’ from virtuous interaction. I expect this is part of the point of being capacious in setting out the models; eventually, we need to get clear on not just the different things we might mean by collective virtues but also what models we need, and in which contexts they apply, in order to help us understand a wide variety of our collective epistemic successes. Additionally, such understandings can be put to work to help us actively design our collaborative teams and communities in ways that can foster strong epistemic performance.

12d Jeroen de Ridder's Response to Commentaries

Thanks to Kate Devitt and Heidi Grasswick for their thoughtful and stimulating comments. They made me see that some of the things I wrote must be reconsidered and that other things can be developed in ways I hadn't considered.

Kate Devitt notes how many structured groups with cognitive goals make extensive use of technology: simple file-storage systems, work management and organization software, advanced monitoring and measurement devices, semiautonomous AI-based solutions that automate part of the group's cognitive labor. Several of the examples I give in my chapter illustrate this, as does Devitt's own chapter in this volume.

Such technological 'scaffolding' can be integrated into all three of my models. In the additive model, technology can support or enhance an individual's cognitive performance so that she comes to possess intellectual virtues she wouldn't have had without technological scaffolding. In science, for example, depositing one's data in a shared repository forces researchers to collect and structure their data with more rigor and carefulness than they otherwise might have. A natural starting point for thinking through this is extant work on extended cognition. The second, interactive, model also has room for technological scaffolding – I already mention this possibility briefly when I discuss interaction with structure in Section 3.3. Technology can correct for human error and compensate vice, as well as enable, promote, or even enforce intellectually virtuous behavior. For the third, emergent, model I unfortunately can do no better than to acknowledge the possibility that technology may give rise to exclusively collective intellectual virtues in both the quantitative and qualitative sense. It certainly seems plausible that a collective with sophisticated technological support can exemplify a degree of intellectual carefulness and accuracy that is unattainable at the individual level. Arguably, the Event Horizon Telescope I mention at the beginning of my chapter is an example. Whether there are also examples of qualitatively novel scaffolded collective virtues remains to be seen.

Devitt's example of a fleet of Tesla cars also raises the fascinating possibility of analyzing complex multi-agent *technological* systems in collective virtue-theoretic terms. This would necessitate recalibrating

the concept of intellectual virtue: while the reliabilist virtue concept – virtues conceived as reliable faculties – might apply fairly straightforwardly to technological systems, it's far from clear that the same goes for responsibilist virtue – acquired intellectual character traits for which one is (partly) responsible and which involve proper motivation. Space is lacking to explore this idea in depth here, but it's an intriguing suggestion that deserves further research.

Devitt worries that my suggestion under the interaction model that properly organized interactions between individuals who lack intellectual virtue might nonetheless produce virtue at the collective level is implausible. That's fair: we certainly shouldn't expect any old interactions between incompetent individuals to produce collective virtue. But even so, incompetence comes in degrees and lacking intellectual virtue does not equal incompetence. The thought behind that version of the interaction model is that cleverly orchestrated interactions between individuals who lack individual virtue might nonetheless produce epistemic excellence at the collective level. This is especially true when these interactions take place in a collective structure or culture that promotes intellectual virtue. Arguably, the Mathematical Tables Project is an example of a case where relatively incompetent individuals nonetheless produce exceptionally reliable outcomes, exactly because they are embedded in a carefully designed system of rules and procedures that eliminates mistakes.

Heidi Grasswick's perceptive remark that I may have run together a weak and a strong version of the additive model is spot on. My take on this is as follows. We can distinguish – at least analytically if not in practice – between (i) a situation where the collectively virtuous outcome is the sum of individually virtuous contributions where the individuals in question would have behaved virtuously regardless of whether they form part of the group (a weak sense), and (ii) one in which the outcome is still the sum of individually virtuous contributions but where the individual virtues are manifested – or manifested to a greater extent – as a result of the presence of other (virtuous) individuals (the strong sense).

Although the strong version of the additive model might now appear very similar to the mere interaction model, there is a crucial difference. The starting point for the additive model is that all group members are individually virtuous. On the mere interaction model, in contrast, the thought is that individually nonvirtuous group members begin to show intellectually virtuous behavior because of their interactions with other group members. I will readily admit that this difference might be very hard to detect in practice: it may not always be clear whether and to what extent an individual possesses intellectual virtues outside the context of a collaborating team. Even so, I believe the distinction is helpful as it shows that individual virtue may not be a necessary condition for a group possessing collective virtue.

Finally, I agree that the relations between the three models deserve further exploration. Grasswick is definitely right that one and the same group might exemplify two or more models at the same time. All individuals on a team may be open-minded, as a result of which the team behaves open-mindedly (additive model). Perhaps those same team members are a bit sloppy individually, but they adhere strictly to the team's standard operating procedures, as a result of which the team's work is meticulous (interactive model). In fact, these procedures might be so rigorously followed that the team as a whole exemplifies a level of meticulousness that no individual could ever attain (emergent model). In other cases, versions of the three models do exclude each other. Grasswick rightly notes that groups with nonvirtuous team members – let alone vicious ones – cannot exemplify the additive model. Similarly, when a group's culture or structure is primarily responsible for its collective virtue (as in the third version of the interaction model), the additive model as well as the other two versions of the interactive model are ruled out. Since the emergent model posits exclusively collective virtues, it is in principle compatible with different distributions of virtues or vices at the level of individual group members and hence with the other two models. Canvassing these interrelations between the models in more detail and in relation to specific virtues (and vices) remains an important follow-up project to what I've done here.

13 Real-Life Collective Epistemic Virtue and Vice

Barend de Rooij and Boudewijn de Bruin

1 Introduction

Moments before Lion Air Flight 610 crashed into the Java Sea with a dramatic loss of life, its pilots frantically searched the jet's handbook in an effort to understand why its nose suddenly pushed down. Working through checklist after checklist amid a growing number of alarms, they never found the information. An official investigation into the causes of the crash later determined that the pilots could never have found what they were looking for. Boeing, which produces the 737 Max jet the pilots were flying, had deliberately omitted crucial information about the flight control system from the manual—a decision that turned out to be part of wide-ranging cost-cutting measures (Komite Nasional Keselamatan Transportasi 2019). Five months after the Lion Air crash, the pilots of an Ethiopian Airlines 737 Max jet saw the nose of their plane unexpectedly push down, and they, too, were unable to find relevant information about the flight control system. Their plane crashed as well.

Even before these two disasters, 737 Max pilots had complained that they “lacked the knowledge” to operate the plane (Fallows 2019). It has since emerged that insufficient documentation is only the tip of the iceberg. The tale of the 737 Max crashes is a story of epistemic vice through and through. At the center of that story is Boeing, the world's largest aerospace company, whose “culture of concealment” is currently the target of an investigation led by the U.S. House Committee on Transportation and Infrastructure (“House Committee” hereafter) (House Committee on Transportation and Infrastructure 2020, 3). In a report of its preliminary findings, the House Committee accuses Boeing not just of failing to provide pilots with proper documentation, but also of withholding “crucial information” from customers and federal regulators (p. 3). Internal communications obtained by the committee further reveal that Boeing ignored several whistleblower complaints and safety warnings issued by its technical pilots and quality assurance officers. Aviation analysts have in fact described Boeing's attitude leading up to the 737 Max crashes as “arrogant” (Gelles et al. 2019, para. 21), and while they do not use the term, from what they write it is clear that this arrogance is at least partly epistemic.

DOI: 10.4324/9780367808952-17

Epistemic arrogance is a canonical epistemic vice: a character trait that obstructs the realization of such epistemic goods as knowledge, understanding, and wisdom.¹ While the concept of epistemic vice has proved useful in analyzing the epistemic state of individuals, philosophers have only recently begun to apply it to *collective* entities such as work teams, boards of directors, committees, or even to entire corporations. We think this move is fully justified by the role collectives play in the economy of knowledge and belief. They are not just the site of beliefs in their own right, as when we say that Boeing *believes* its planes to be safe. More than that, they are key in the *transmission* of beliefs, as when a pilot relies on Boeing for proper documentation.

But what does it mean to say that Boeing was *arrogant*?

This chapter critically examines extant theorizations of collective epistemic virtue and vice. It lays out certain conceptual problems and proposes ways of overcoming these problems. We argue for what could be called a *functionalist* account according to which epistemically virtuous groups are organized so as to function *as* epistemically virtuous agents. On the view we defend, an epistemically virtuous organization typically has three components: it exhibits *organizational support for virtue*; it has *organizational remedies against vice* in place; and it *matches the individual virtues of employees to the organization's functions*, for instance in hiring decisions. Organizations may manifest collective epistemic vice if they fail to enact a corporate structure that is virtuous in this way. One key aim of this chapter is to illustrate the practical *real-life relevance* of an approach to collective virtue epistemology, which is also conceptually and empirically sound. We therefore spend considerable time investigating Boeing's *epistemic corporate culture* (de Bruin 2020).

2 Collective Epistemic Virtue and Vice

Epistemic virtues are features that make us excellent *qua* producers and consumers of epistemic goods, such as knowledge, understanding, or wisdom. By contrast, epistemic vices obstruct the realization of these epistemic goods. Virtue epistemologists disagree somewhat over the nature of these features. For virtue reliabilists such as Ernest Sosa (2007) and John Greco (2010), epistemic virtues comprise *all* stable dispositions that reliably produce true beliefs. Prime examples of reliabilist virtues include such cognitive faculties as sense perception and reliable memory. Virtue responsibilists such as Lorraine Code (1987) and Linda Zagzebski (1996), on the other hand, characterize epistemic virtues primarily as the *character traits* that mark an excellent knower. In this picture, an epistemically virtuous knower not only reliably forms true beliefs, but also cultivates such epistemically virtuous character traits as honesty, open-mindedness, and intellectual courage. Cultivating these traits requires that we are moved by virtuous epistemic *motives*, such as love of wisdom.

While the literature on epistemic vices has only recently emerged (see, e.g., Baehr 2010; Battaly 2014, 2016; Cassam 2016, 2019; de Bruin 2015), they are typically conceived of as the inverse of epistemic virtues. Thus, epistemic vices may include such unreliable faculties as poor vision and obstructive character traits as closed-mindedness, overconfidence, or hubris.

We not only attribute epistemic virtues and vices to individuals, we also regularly attribute them to groups. We commend, for instance, the International Consortium of Investigative Journalists (ICIJ) for its display of intellectual courage, or we reproach the financial service providers whose misconduct the consortium unmasked as careless and dishonest. As intuitive as we find it to use the language of virtue and vice to talk about groups, the metaphysical status of these attributions is far from straightforward, though. Are we merely using a linguistic shortcut to talk about the features of its members, or do we say that the group exhibits these features *qua* group?

Summativists are poised to answer that group features reduce to individual features, and so that groups lack these features as subjects in their own right.² On a basic summativist analysis, group *G* exhibits virtue or vice *V* only if a sufficient number of its members exhibit *V*. If the ICIJ consortium is a courageous institution, summativists maintain, this is because its *individual journalists* display the virtue of courage (in speaking truth to power, say). Since this entails that only individuals can be the proper subject of virtues and vices, summativists hold that the most we can do when we attribute these traits to groups is make summary reference to the traits of its individual members. We would be mistaken if we believed, for instance, that the ICIJ is the seat of courage as a subject in and of itself.

Summativism enjoys a great deal of initial plausibility. It meshes well with the widespread conviction that individual agents are the basic explanatory units of all social phenomena (the doctrine of *methodological individualism*). Moreover, it is clearly correct as an account of at least *some* collective virtues; for when we praise our group of students for their diligent work ethic, we really do appear to praise the character of our individual students.

Yet summativism often oversimplifies the relation between a group and its members. Consider Reza Lahroodi's (2007) example of a group that is collectively narrow-minded even though it is for the most part composed of individually open-minded members. We could think of the board of directors of an aerospace company. As individuals, these board members are open-minded about such things as aerospace innovation, and they are disposed to give reasonable innovative ideas a fair hearing. Collectively, however, the board is not so disposed. The board believes the company's market position is sufficiently secure not to invest in innovation, and so when innovation comes up as a topic in the boardroom, it often dismisses these ideas without giving them fair consideration.

Lahroodi's case is construed in such a way that if we were to tally the number of individually narrow-minded directors, we would come up empty. He observes that a summativist should conclude that the board is not collectively narrow-minded either (i.e. that the board is not narrow-minded *qua* group). This, he claims, seems wrong, as the board routinely rejects innovative ideas out of hand. Lahroodi therefore contends that summativism is incorrect as a general account of group epistemic virtue and vice: we cannot always analyze such traits as mere sums of the traits possessed by individuals.³

Perhaps a sophisticated summativist may be able to account for Lahroodi's example, for perhaps it is not individual *virtues* we should tally, but other individual features. Yet a growing number of philosophers take examples of Lahroodi's kind to motivate the search for *nonsummativist* accounts of collective virtue and vice (e.g., Fricker 2010). Nonsummativists claim that the members of a group sometimes interact in such a way that they form collective agents whose properties are distinct from the properties exhibited by these members themselves.⁴ In other words, they hold that groups can be more than, or at least different from, the sum of their parts, as Lahroodi's example illustrates.

A leading nonsummativist account of group agency is due to Margaret Gilbert (1989, 2013). According to Gilbert, some groups form what she calls *plural subjects*, with intentions, beliefs, and other agential features of their own. These plural subjects are instantiated when two or more individuals jointly and openly commit to upholding these features *as a body* (Gilbert 2013, 32). The journalist members of the ICIJ, for instance, form a plural subject of the intention to uncover fraud to the extent that they jointly and openly commit to investigate fraud and money laundering as a body, or as one. Gilbert seems to intend her notion of doing something as a body, or as one, to be read metaphorically, as she does not believe that plural subjects *literally* possess a body of their own. What she thinks is that the parties to a plural subject coordinate their actions so as to emulate a single body; hence the spirit of methodological individualism is preserved.

Gilbert (2013) is clear that plural subjects are irreducibly collective entities because the constitutive joint commitments do not reduce to *personal* commitments. This opens up a logical space in which these commitments diverge: you can be jointly committed to narrow-minded practices, even if you are personally committed to being open-minded, just as in our example. The technical details of Gilbert's argument need not detain us here, but it may be helpful to point out that the difference between these two types of commitment is brought out by normative expectations accompanying them. When you personally commit to something, you can unilaterally rescind the commitment whenever you like. You do not owe it to anyone to follow through on your commitment. But if you committed to something jointly with others, you and the other

members of a resulting plural subject incur obligations towards each other. No member has the standing to rescind these joint commitments unilaterally; they can only be rescinded (without violating social norms) if everyone agrees. Joint commitments are, that is, intrinsically other-involving, and this is why we cannot perform a summative reduction of these commitments to personal ones.

While Gilbert's account can help us understand how groups could instantiate properties their members lack individually, Lahroodi doubts that it provides a fully viable model for collective virtue and vice. The problem, he claims, lies in Gilbert's requirement that joint commitments be *open*, or *transparent*, to all parties involved. Gilbert indeed holds that one of the prerequisites for being jointly committed to something is that the members of the plural subject have expressed to each other their willingness to be so committed, thereby signaling acceptance of the content of the commitment in question. As Lahroodi argues, however, this lacks plausibility when it comes to virtue and vice. A group can be open-minded, he thinks, even if its members do not know that it is open-minded, let alone have openly committed to open-mindedness; what matters is only its disposition to give a fair hearing to contrary ideas.⁵

There are various ways of responding to Lahroodi's concerns. Miranda Fricker (2010), for instance, argues that there is no special philosophical puzzle in holding that plural subjects can manifest epistemic virtues or vices none of its members are aware of. They may simply fail to know that the traits to which they have jointly committed count as virtuous or vicious. Just as some individuals manifest the virtue of modesty without knowing that they could be adequately described as *modest*, so the members of a plural subject may be jointly committed to routines, values, or procedures without knowing that these features constitute virtuous or vicious traits. The directors of our aviation company, for instance, need not be aware of having committed to *narrow-mindedness*, under that description. Our version of Lahroodi's example is more plausibly construed as involving a positive commitment to maintaining the corporation's legacy, which happens to have the unhappy consequence of reducing investments in research and development.⁶

Still, various problems with a plural subject approach remain.⁷ For one, it is unclear whether Fricker's response will satisfy critics such as Lahroodi. While Fricker is arguably correct that the members of a plural subject need not construe the trait they have jointly committed to as virtuous or vicious for it to have the relevant virtue or vice, they do, on a Gilbertian analysis, need to have somehow expressed a willingness to be committed to the trait in question. You might think that even this requirement is too strong. In Lahroodi's words, it simply does not seem totally felicitous to claim that members "have to jointly accept to exercise a trait for the group to have a trait" (p. 292); there are many groups that, on the face of it, exhibit traits that their members have not accepted

to exercise. Negligence may be a case in point, if it is thought that a reluctance to accept *any* commitments toward safety and diligence makes a group negligent.

Another issue concerns the empirical adequacy of a plural subject approach. We often attribute virtues and vices to universities, multinational corporations, NGOs, and other large collectives that may be composed of thousands of members, many of whom will never interact. Boeing, for instance, has over 150,000 employees across 65 countries. It has three business divisions, and dozens of offices and manufacturing plants. Supposing that Boeing suffers the vice of arrogance, as some analysts maintain, is it really plausible to claim that its employees have all jointly expressed to each other a readiness to commit to arrogant qualities?

In light of these concerns, we may want to turn to an alternative non-summative analysis of group agency. We suggest that a *functionalist* analysis of group agency provides a way of modeling collective epistemic virtues and vices that is particularly promising if you are driven by real-life practical concerns. On a functionalist analysis, groups possess agency insofar as they are systems that *function* as agents.⁸ Christian List and Philip Pettit (2011) illustrate this using a classic belief–desire model of agency. An agent, List and Pettit argue, is a system that exhibits three features: it has beliefs about what the world is like; it has desires as to how the world should be, and it has the capacity to act on these desires on the basis of these beliefs. Individual human persons satisfy these conditions, but so do many other systems, including robots, animals, and, List and Pettit maintain, some groups. A business organization, for instance, typically has desires (say, to maximize profits), beliefs (say, about market conditions), and the capacity to produce and sell goods or services on the market in order to realize these desires.

According to List and Pettit, the relationship between a group and its members is one of supervenience, so there cannot be a difference *qua* group-level beliefs, desires, and other features without there being a difference *qua* features possessed by individual group members (taking the procedure by means of which these individual features are *aggregated* into account). Accordingly, functionalism does not militate against methodological individualism. Crucially, List and Pettit argue that the members of a group can realize group agency in various configurations. There are many conceivable aggregation functions taking us from individual to group beliefs and desires, and numerous ways in which groups could act on these beliefs and desires. A group could use majoritarian voting methods, for instance, but it could also choose to adopt a dictatorial Chief Executive Officer (CEO).⁹

The crux of List and Pettit's nonsummative analysis lies in the multiple realizability of group-level features. On a functionalist analysis, the same group beliefs and desires can be produced by different aggregation

functions and/or on the basis of different individual *input* beliefs and desires. When a corporation fires one employee, for instance, these input beliefs and desires marginally change, but that change need not translate into changes at the group level. Since different aggregation functions and inputs to these functions can produce the same group-level results, List and Pettit conclude that it is frequently impossible to perform a summative reduction of group-level features to individual features. In such cases, groups are the bearer of their features as subjects in their own right.

If this shows that groups can function as agents, the question remains whether they can function as epistemically virtuous or vicious agents. In a response to Lahroodi, Todd Jones (2007) answers affirmatively.¹⁰ Functionalists maintain that groups can be organized to realize various cognitive processes (including belief-forming processes). But “[i]f groups can have cognitive processes,” Jones writes, then “they can have better and worse working cognitive processes and thus exhibit epistemic virtues” (p. 441). Indeed, once we view groups as functional kinds, we must conclude that “there are many different ways for groups to have epistemic virtues” (p. 447). And, we might add, epistemic *vices* too. This is because there are many different ways in which groups can be organized to implement cognitive functions. Of the many different conceivable aggregation functions, which take us from individual to group attitudes, that List and Pettit (2011) allow for, some are conducive to the group’s epistemic ends, while others obstruct it. The same goes for the decision procedures the group uses to translate these attitudes into action.

The challenge, then, is to identify those organizational structures within which group members combine so as to function as an agent that exhibits a collective epistemic virtue or vice. Despite the work by Jones, this is still largely an open task. In the section that follows, we build on earlier work by De Bruin (2015) and take a closer look at the epistemic misconduct disaster at Boeing.¹¹

3 Case Study: The Boeing 737 Max Disaster

As we suggested in the previous section, organizations can function as epistemically virtuous or vicious agents in many different ways. One reason for this is that organizations can exhibit a wide variety of what Peter French (1979) calls *corporate internal decision structures* (CIDs). CIDs comprise *responsibility flowcharts* that determine the hierarchical relationships between the organization’s members, and *corporate decision recognition rules* that determine the mechanisms by which corporate decisions are made. Often anchored in corporate charters, articles of association, and other official documents, these flowcharts and recognition rules assign particular roles to the members of an organization and determine the rights and duties associated with these roles.

For example, CIDs determine the conditions under which someone is authorized to speak on behalf of the organization, but also fix internal hierarchies and how beliefs and decisions are aggregated within the organization.

De Bruin (2015) has argued that CIDs are important loci of corporate epistemic virtue or vice. An epistemically virtuous organization, on this view, is structured such that its responsibility flowchart and corporate decision recognition rules together produce a tendency towards epistemically virtuous behavior and against vice. The CID of an open-minded organization, for instance, will dispose that organization towards taking contrary ideas seriously. For this to work optimally, organizations must satisfy three conditions. First, epistemically virtuous organizations must ensure that group members exhibit the virtues required by their roles within the organization (*virtue-to-function matching*). Secondly, these organizations must encourage the exercise of these virtues by providing a supportive environment (*organizational support for virtue*). And finally, epistemically virtuous organizations must include safeguards against epistemic vice (*organizational remedies against vice*). While epistemically virtuous organizations typically realize all three of these conditions, a failure to realize any one of them can obstruct the epistemic ends of an organization and thus produce epistemically vicious behavior.¹² To demonstrate the practical relevance of our approach when it comes to understanding epistemically vicious behavior in collectives, we now turn to the recent case of Boeing.

4 Background

Boeing is the largest aerospace company in the world, producing commercial and military airplanes as well as rockets, satellites, and security and defense systems. Founded in 1916, it has an impressive track record. Its bombers played a crucial role in deciding World War II; its 747 Jumbo Jet revolutionized the mass tourist industry; and its contributions to space travel include NASA's first probe to circle the moon as well as part of the rockets NASA later used to land astronauts on the moon.

Boeing's most famous accomplishment is, however, the Boeing 737. It was the best-selling jet in aviation history, until two crashes of its newest-generation model, the 737 Max, killed 346 people in 2018 and 2019. At the time of writing this chapter, Boeing's chief rival—the European conglomerate Airbus—has overtaken sales of the 737 with its A320, and sector analysts predict that the 737 is unlikely to catch up anytime soon.

Some historical and technical background is important. Boeing launched the first two generations of 737 jets in the 1960s and 1980s, and faced little competition from rival manufacturers until Airbus introduced its A320 in 1987. By the 1990s, it became clear that Boeing had a

problem on its hands, as many of its long-time clients showed significant interest in Airbus. To remain competitive, Boeing therefore introduced a third-generation 737, the 737 Next Generation (NG), with greater fuel capacity, an updated cockpit, and more seats—ten years after Airbus's A320. As it turned out, however, A320 sales far surpassed sales of the 737NG, and in the mid-2000s, analysts therefore believed that Boeing should make a more radical move and design an entirely new aircraft (Thomas 2006).

Boeing postponed decision-making on whether or not to design a new jet for years, and in 2010 it was again Airbus that made the first move. Airbus decided against developing a new plane, but chose to refit the A320 with more fuel-efficient engines. This practice is known as *re-engining*.

Under the impression that Airbus had misread the markets, Boeing dismissed the viability of re-engining. The head of Boeing's commercial airplanes division thought that Airbus's re-engining was financially unwise, and would lead to "a plane that carriers didn't really want," and so, he thought, Boeing "could wait until the end of the decade to produce a new plane from scratch" (Gelles et al. 2019, para. 13). Boeing's then-CEO James McNerney likewise stated that "the leader in the clubhouse is the all-new aeroplane" (Weitzman 2011, para. 5).

Boeing was entirely wrong. Oil prices were surging, and carriers did want more fuel-efficient engines, and they wanted them fast. So when one of Boeing's biggest clients, American Airlines, announced in 2011 that it would move part of its business to Airbus, Boeing was forced to reverse course (Odell 2011). Yet by that point, Boeing's ignoring evidence about consumer preferences had cost it precious time. While Airbus had been successful in re-engining the A320, Boeing's 737 Max suffered, as we saw, two dramatic crashes shortly after entering service in 2017. It is important to understand what happened from a technical perspective.

Unlike its predecessors, the engine of the 737 Max is attached forward on the wings rather than suspended under the wings. This forward engine placement creates particular aerodynamic challenges. With too much power to the engine, the plane's nose may go up, increasing the so-called *angle of attack* (AOA), which is the angle between the wing and the flow of air. A certain AOA is needed to lift the plane, but if a plane exceeds the optimum AOA its lift will suddenly decrease because the air no longer flows smoothly along the wings but becomes turbulent, a condition called *stall*.

Stalling is dangerous. If your paper airplane goes up too steeply, it does not get very far but falls, like a stone, and crashes. Since any aircraft is at risk of stalling, commercial aircraft have sophisticated stall control systems in place. Although the technical details of the 737 Max's stall-prevention systems are still under investigation, a key component

seems to be a software package called the Maneuvering Characteristics Augmentation System (MCAS), which receives information from an AOA sensor close to the jet's nose, and pushes the nose down when the critical AOA is exceeded. Flight data indicate that MCAS received false input from the AOA sensor. It wrongly suggested the plane was about to stall, and therefore automatically pushed the nose down, crashing the plane (House Committee on Transportation and Infrastructure 2020). AOA sensors are prone to malfunction, exposed as they are to low temperatures and lightning, and they are sometimes installed incorrectly. But Boeing made its stall-prevention system rely on only one AOA sensor in the re-engined 737.

5 Virtue-to-Function Matching

At the moment of writing this chapter, there seems to be a fair degree of consensus among experts suggesting that the decision to rely on one AOA was a key—and human—error explaining the two crashes. But who took that decision? Some observers have said the 737 Max was “designed by clowns who are in turn supervised by monkeys” (Bushey and Stacey 2020). This brings us to the first element of corporate virtue: virtue-to-function matching. Any organization has numerous goals. Boeing has the ambition to be the primary choice among pilots. One of its slogans was: “If it's not Boeing, I'm not going.” It wants to develop first-rate technology. It wants to maximize profits for its shareholders, and many other things. Achieving these goals involves accomplishing a wide variety of tasks. The design and construction of a wing, for example, requires modeling the aerodynamics of the wing and applying the materials science behind the composites involved in building the wing. It involves the know-how of technicians who assemble the wing, and the expertise of testers who determine whether the wing is safe and functions as intended. It also requires patent lawyers who scrutinize the project for any liabilities. Moreover, before the project even starts, accountants must draft budgets.

In technology-intensive industries, most jobs have substantial epistemic components. Knowledge (information) must be collected, engineered, stored, processed, evaluated, shared with colleagues, communicated to the workforce, and so on. The thought behind virtue-to-function matching as the first component of an epistemically virtuous organization is that these different types of epistemic work are facilitated by different epistemic virtues. The virtues of curiosity and wonder facilitate new insights through experimentation, engineering, modeling, and reflection. Humility and temperance help evaluate the relevance and reliability of new information. Sharing and communicating information is an exercise in epistemic generosity. And storing information requires attentiveness and care.

No organization can hope to find members that score high on each of these epistemic virtues. Some persons are curious and creative, others more attentive or generous. To reliably achieve the various epistemic ends of an organization, management should therefore ensure that the members of that organization have the epistemic virtues required by their roles and responsibilities within the organization. While this may be difficult to contest conceptually, empirical practice is often complicated.¹³

The most widely studied function in business scholarship is that of the managing director or CEO of a firm. The CEO is the firm's key representative vis-à-vis its owners (shareholders) and other stakeholders, and must have a clear view of the firm's long-term strategy. The CEO is the first and foremost decision maker of the firm and bears the main responsibility for its decisions.

So let us turn to James McNerney, at the helm of Boeing between 2005 and 2015. A day after McNerney announced his departure at Boeing, the prominent aerospace industry analyst Richard Aboulafia (2015) contributed an insightful profile to the respected American business biweekly *Forbes*. Although Aboulafia opens with the observation that McNerney pleased Boeing's shareholders, he reprimands him for leaving a "toxic legacy" (para. 2). Aboulafia details how McNerney's concern for shareholder interests led him to move production to new facilities, and cut pensions and salaries: "Taking away pensions at a time of record sales is a terrible way to motivate workers to go the extra mile" (para. 6).

The question to ask is whether McNerney's epistemic character traits matched his function as a CEO. Aboulafia thinks not. McNerney had no prior experience with aviation when he started at Boeing. His previous job was CEO at 3M (think Post-Its and face masks). But as Aboulafia says: "If a CEO comes from a different industry and doesn't try to learn what makes aviation distinct, he's likely to apply a one-size-fits-all template" (para. 8). And indeed, like many American companies, 3M had faced increasingly intense competition from low-cost countries, which arguably justified the drastic methods McNerney had deployed to ensure the firm's survival. But aviation is a very different industry, with only two major global players and precious little competition from outsiders. In such a market, Aboulafia says, "[a]n *experienced* and *motivated* workforce ... is the most important asset a company has" (para. 10, emphasis ours).

Several commentators do indeed implicate an *inexperienced* and *unmotivated* workforce in the safety lapses at Boeing's manufacturing plants and its poor handling of the crashes (Gelles 2020; Kitroeff and Gelles 2019). Yet we want to be cautious and avoid suggesting a direct link between suboptimal virtue-to-function matching—McNerney's lack of curiosity about the aviation industry—and the 737 Max disasters. We believe that corporate epistemic virtue requires more than that.

6 Organizational Support for Virtue

Besides ensuring virtue-to-function matching, an epistemically virtuous organization should strive to create and maintain an environment that is sufficiently conducive to epistemic virtue in which employees should, among others, feel free to ask questions, share knowledge, criticize each other, and investigate things. In such an environment, senior employees will have to pay attention to what juniors say, openly acknowledge the value of their input, and should not be above changing their minds on the basis of this input. In short, epistemically virtuous organizations should have a system of incentives (in the broadest sense of the word) in place to stimulate and support epistemically virtuous behavior.

The earlier claim that the 737 Max was “designed by clowns who are in turn supervised by monkeys” is surely hyperbole. But reports in the media and official investigative findings provide ample evidence that the 737 Max was designed and produced in a decidedly suboptimal epistemic environment. For example, it appears that the commercial pressures at Boeing obstructed the creativity and innovativeness of its engineers. Here are some examples from the congressional hearings and media reports. Engineers were requested to deliver technical drawings at “double the normal pace” (para. 8), and “sloppy blueprints” (para. 29) were delivered by “rushed designers” (para. 29; Gelles et al. 2019). Engineers were forced to make as few changes as possible to the aircraft so as to minimize the need for new pilot training (as this would make the plane less attractive to prospective buyers who would have to pay for the training). This was felt as considerably frustrating their creativity: “there was so much opportunity to make big jumps, but the training differences held us back” (ibid., para. 38). Rather than harnessing the virtues of its engineers, Boeing held them back.

Information sharing was minimal and discouraged throughout design and production processes. When the plane was finally constructed, Boeing was highly reluctant to share information with pilots, who report not understanding particular signals, and finding no relevant explanation in-flight manuals (Fallows 2019). Even prior to the crashes, pilots complained about “[p]oor training and even poorer documentation” (para. 52) and a lack of information about “the highly complex systems that differentiate [the 737 Max] from prior models” (para. 64), with the result that they “lacked the knowledge” (para. 84) required to fly the plane safely (Fallows 2019). In fact, information about MCAS, the software system implicated in both crashes, was missing from the manuals altogether. Boeing reasoned that since MCAS would operate “in the background” (para. 47) pilots would not need to be briefed on it (Gelles et al. 2019).

To be sure, organizational support for epistemic virtue does not require that everyone knows everything. You do not need to understand

Linux code to use the operating system responsibly. But in this case, we are talking about *pilots*, who were confronted with alarm signals their documentation failed to explain to them *during a flight*.

The House Committee called this a “culture of concealment” (House Committee on Transportation and Infrastructure 2020, 3). This culture also manifests itself in other areas. A key form of organizational support for epistemic virtue is that people can speak freely, without fear of repercussions.¹⁴ Only if employees can be confident that they can talk to superiors without the risk of losing their jobs or being relocated will they speak frankly. Instead of encouraging its employees to report on safety issues, Boeing swept safety issues under the rug and concealed them from regulators. Boeing employees did bring forward various whistleblower complaints to the effect that superiors actively discouraged them from reporting manufacturing errors and other safety violations. But they were discouraged from doing so, and some faced retaliation when they did. Quality managers who noticed that defective parts were installed in planes were told not to worry and removed from projects if they persisted.

Perhaps Boeing’s most lamentable decision was making its MCAS flight control system rely on a single AOA sensor, as mentioned earlier. Employees expressed doubts about “whether the system was vulnerable to malfunctioning if a single sensor failed” as early as 2015 (Gelles and Kitroeff 2019, para. 6). For reasons that are not entirely obvious, their concern received no uptake, although it seems likely that Boeing underplayed the danger of an MCAS failure to ease certification procedures (Gates 2019). While these procedures should have served as a check on the adverse epistemic conditions at Boeing, preliminary investigative findings suggest they failed to act as an effective remedy against Boeing’s epistemically vicious tendencies. This brings us to the last element of organizational epistemic virtue.

7 Organizational Remedies Against Vice

Organizations must offer supportive environments to enable epistemically virtuous individuals. Most of us are no virtue epistemic superheroes, though, and hence organizations must also have remedies in place that mitigate the effects of epistemic vice. These remedies can be implemented at various levels within an organization. An illustrative example of a *macro-level* remedy shows how organizations can protect themselves from adopting one-sided or biased views. An employee or team has invested considerable time and resources in developing a plan for a new product, and presents it to a decision maker within the organization—the “boss.” What will the boss do? In many organizations, bosses decide by themselves, and give the project the green light, or not.

Looks good? Not from a virtue epistemological point of view. A procedure like this is asking for corporate narrow-mindedness. At the level of the corporation, only one side of the story is listened to: the story that puts the project in favorable light. The boss could do much better by asking some person or team with no stakes in the project to come up with as many arguments against the project as possible, and then, with the pros and cons in hand, decide.

Organizing opposition or *dissent* is an essential macro-level remedy against various epistemic vices. There is always a risk of being overconfident about a project you are invested in, of rushing to conclusions, or of narrow-mindedly ignoring evidence that suggests a more downbeat view of your plan's prospects. There will always be team members who do not share relevant information as extensively as necessary. People make mistakes, are forgetful, and succumb to sunk-cost fallacies and continue working on a project even after they see that it is not really worth the investment any longer. To mitigate these and other biases, organizations have to develop remedies such as organizing dissent.

One form of dissent that lies at the heart of the aviation industry centers on independent governmental bodies regulating the industry. In the United States that mandate falls on the Federal Aviation Administration (FAA), whose setup and underlying rationale resemble the Securities and Exchange Commission and the Food and Drug Administration. Boeing employees are in close contact with the FAA at all times. There are good reasons for this: designing an airplane is costly, so you do not want to go through the entire design process only to learn that the FAA refuses certification. But the preliminary House Committee report suggests that Boeing and the FAA may have gotten much too close. A central point of concern is to do with *authorized representatives*. These are people employed and paid by Boeing, but tasked to represent the interests of the FAA. Email and WhatsApp conversations show that these representatives nudged the FAA into accepting the view that MCAS, the flight control system, would merely be a "speed trim function," not requiring additional certification and pilot training (House Committee on Transportation and Infrastructure 2020, 3, n. 16). The FAA agreed.

Who pays the piper calls the tune? Whenever sharing your knowledge comes at a cost to yourself or your employer, conflicts of interest are likely to arise. Authorized representatives are not alone here, witness elaborate codes of conduct managing conflicts of interests in health care, financial services, accountancy, engineering, and many other professions. If effective remedies are in place that guarantee objectivity and epistemic independence, many of these potential conflicts can be averted, and authorized representatives were indeed shaped into one of Boeing's key devices to organize dissent. They were, that is, a key remedy against epistemic vice.

The House Committee report strongly suggests, however, that this remedy dramatically failed. Not only did authorized representatives misconstrue the flight control system to the FAA, they also failed to inform the FAA of various safety concerns. For instance, they did not warn the FAA that Boeing sold aircraft with inoperative devices meant to detect AOA discrepancies, although that problem was known internally as early as 2015. When Boeing finally set about fixing this fault in its AOA indicator software in 2017, an authorized representative signed off on Boeing's plan to postpone the required software update to 2020, again failing to inform the FAA. And perhaps most damningly, they concealed crucial safety information during the plane's development. One authorized representative questioned the safety of relying on a single AOA sensor in internal communications, but that concern was brushed aside and not reported to the FAA. Moreover, it turns out that several authorized representatives were aware of a Boeing analysis showing that pilots had at most 10 seconds to respond to unusual signs from the flight control system, and that a failure to act accordingly could be "catastrophic" (House Committee on Transportation and Infrastructure 2020, 3). But they never shared this knowledge with the FAA. At these and other junctures, Boeing's authorized representatives could have counteracted the epistemic misconduct that was generated by the commercial pressures under which Boeing was producing the 737 Max. By failing to do so, they instead let vice run rampant.

8 Conclusion

In this chapter, we have critically examined extant conceptualizations of collective epistemic vice and virtue, and we have defended our own, functionalist account. Following this approach, collective vice and virtue are instantiated when groups are organized so as to *function* as an epistemically virtuous or vicious agent. While the ways in which group agents can enact virtuous or vicious corporate structures no doubt vary, we have singled out three elements of such structures: virtuous organizations ensure that group members have the epistemic virtues required by their role within the organization (*virtue-to-function matching*); they provide *organizational support for these virtues*; and they enact *remedies against epistemic vice*.

Correspondingly, organizations can collapse into an epistemic vice if they fail to enact a corporate structure that is virtuous in this way. In order to illustrate this, we presented a case study of Boeing's epistemic conduct surrounding the crashes of two 737 Max jets in 2018 and 2019. We showed how Boeing's leadership appears to have lacked some of the virtues required of corporate decision makers; how commercial pressures generated an environment that was not conducive to epistemic virtue; and how Boeing's remedies against vice failed to offset these pressures.

None of this is to say that the two deadly crashes are entirely to blame on epistemic problems. But collective epistemic vice undoubtedly played a part.

Acknowledgments

Research for this chapter was funded by the Dutch Research Council (NWO) grant number 360–20–380.

We owe warmest thanks to the editors of this volume, and to Miranda Fricker, Frank Hindriks, and Jeroen de Ridder for detailed comments on an earlier version of this chapter.

Notes

- 1 We do not mean to take a firm position on the nature of epistemic vice in this chapter. That is why, for present purposes, we have deliberately selected this rather broad characterization of epistemic vice that is loosely based on Cassam's (2019) definition of epistemic vice.
- 2 The term *summativism* is due to Quinton (1975). See Gilbert (1989) for discussion. Fricker (2010), Lackey (2016), and Gilbert (1989) prefer the term *summativism*. Lahroodi (2007) uses *individualism*, and List and Pettit (2011), *eliminativism*, all with subtle distinctions. We use *summativism* without privileging any of the extant views.
- 3 Lahroodi's case amounts to a virtue epistemological version of what Lackey (2016) calls *divergence arguments* in her discussion of group justification.
- 4 Lahroodi (2007) uses the term *anticorrelativism* instead of *nonsummativism*. List and Pettit (2011) prefer *realism about group agency*, with subtle distinctions.
- 5 Driver (2001) individuates a class of *vices of ignorance* that we are necessarily unaware of having (if we have them). Cassam (2019) similarly identifies a range of *stealthy vices*.
- 6 Our construal of Lahroodi's case is meant to be plausible. Anticipating our later discussion: Boeing has reportedly resisted fully embracing computerized flight control technology for decades, believing pilots prefer to be in charge at all times. Its main competitor Airbus has installed extensive computer technology in aircraft since the early 1980s (Gelles et al., 2019).
- 7 For a critical discussion of Fricker, see Byerly and Byerly (2016), Cordell (2017), and Konzelmann Ziv (2012).
- 8 Many functionalists believe that functioning as an agent simply is what it means to be an agent. Similarly, we believe that groups that function in a virtuous or vicious way really do have group-level virtues or vices—they are not simply *as-if* virtues or vices.
- 9 See de Bruin (2018) for an analysis of aggregating quantitative, financial judgments in the boardroom.
- 10 Byerly and Byerly (2016) also appeal to a functionalist analysis when they argue that corporate virtues and vices are multiply realizable: a group can replace one or more of its members without thereby losing its virtues or vices.
- 11 We depart from existing work on corporate virtue that applies virtue theoretical insights to organizational practice, but focuses on moral virtue (see, e.g., Gowri, 2007; Moore, 2005, 2015; Sandin, 2007). Earlier virtue

- epistemological work in the context of business covered *individual* epistemic virtues (de Bruin, 2013).
- 12 It is not our ambition to provide a conceptual analysis of corporate virtue. As such, we remain agnostic about the necessary and sufficient conditions for corporate virtue. However, the factors we identify—the presence of *virtue-to-function matching*, *organizational support for virtue*, and *remedies against epistemic vice*—tend to be sufficient for corporate virtue in typical real life cases.
 - 13 Further, it may not always be easy to link specific tasks to corresponding virtues. We follow Jason Baehr's (2011, p. 21) meticulously argued taxonomy here.
 - 14 Fricker (2020) discusses a similarly vicious organizational culture, at the BBC.

References

- Aboulafia, R. (2015, June 24). Boeing Will Pay High Price for McNerney's Mistake of Treating Aviation Like It Was Any Other Industry. *Forbes*. Retrieved from <https://www.forbes.com/sites/richardaboulafia/2015/06/24/boeing-mcnerney-and-the-high-price-of-treating-aircraft-like-it-was-any-other-industry/>
- Baehr, J. (2010). Epistemic Malevolence. *Metaphilosophy*, 41, 189–213. <https://doi.org/10.1111/j.1467-9973.2009.01623.x>
- Baehr, J. (2011). *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford: Oxford University Press.
- Battaly, H. (2014). Varieties of Epistemic Vice. In J. Matheson & R. Vitz (Eds.), *The Ethics of Belief*. Oxford: Oxford University Press.
- Battaly, H. (2016). Epistemic Virtue and Vice: Reliabilism, Responsibilism, and Personalism. In *Moral and Intellectual Virtues in Western and Chinese Philosophy: The Turn toward Virtue* (pp. 99–120). New York: Routledge.
- Bushey, C., & Stacey, K. (2020, January 10). Boeing Workers Mocked Regulators Over 737 Max Approval. *The Financial Times*. Retrieved from <https://www.ft.com/content/7785de46-3350-11ea-9703-eea0cae3f0de>
- Byerly, T. R., & Byerly, M. (2016). Collective Virtue. *Journal of Value Inquiry*, 50, 33–50. doi:10.1007/s10790-015-9484-y
- Cassam, Q. (2016). Vice Epistemology. *The Monist*, 99(2), 159–180. doi:10.1093/monist/onv034
- Cassam, Q. (2019). *Vices of the Mind: From the Intellectual to the Political*. Oxford: Oxford University Press.
- Code, L. (1987). *Epistemic Responsibility*. Hanover, NH: University Press of New England.
- Cordell, S. (2017). Group Virtues: No Great Leap Forward with Collectivism. *Res Publica*, 23(1), 43–59. doi:10.1007/s11158-015-9317-7
- de Bruin, B. (2013). Epistemic Virtues in Business. *Journal of Business Ethics*, 113(4), 583–595. doi:10.1007/s10551-013-1677-3
- de Bruin, B. (2015). *Ethics and the Global Financial Crisis: Why Incompetence Is Worse than Greed*. Cambridge: Cambridge University Press.
- de Bruin, B. (2018). Moral Responsibility for Large-Scale Events: The Difference between Climate Change and Economic Crises. *Midwest Studies in Philosophy*, 42(1), 191–212. doi:10.1111/misp.12090

- de Bruin, B. (2020). Epistemic Corporate Culture: Knowledge, Common Knowledge, and Professional Oaths. *Seattle Law Review*, 43(2), 807–839.
- Driver, J. (2001). *Uneasy Virtue*. Cambridge: Cambridge University Press.
- Fallows, J. (2019, March 13). Here's What Was on the Record about Problems with the 737 Max. *The Atlantic*. Retrieved from <https://www.theatlantic.com/notes/2019/03/heres-what-was-on-the-record-about-problems-with-the-737-max/584791/>
- French, P. (1979). The Corporation as a Moral Person. *American Philosophical Quarterly* 17(3): 207–215.
- Fricker, M. (2010). Can There Be Institutional Virtues? In T. Gendler & J. Hawthorne (Eds.), *Oxford Studies in Epistemology* (Vol. 3, pp. 235–252). Oxford: Oxford University Press.
- Fricker, M. (2020). Institutional Epistemic Vices: The Case of Inferential Inertia. In I. J. Kidd, H. Battaly, & Q. Cassam (Eds.), *Vice Epistemology*. Milton: Taylor & Francis.
- Gates, D. (2019, March 17). Flawed Analysis, Failed Oversight: How Boeing, FAA Certified the Suspect 737 MAX Flight Control System *Seattle Times*. Retrieved from <https://www.seattletimes.com/business/boeing-aerospace/failed-certification-faa-missed-safety-issues-in-the-737-max-system-implicated-in-the-lion-air-crash/>
- Gelles, D. (2020, January 10). 'I Honestly Don't Trust Many People at Boeing': A Broken Culture Exposed. *The New York Times*. Retrieved from <https://www.nytimes.com/2020/01/10/business/boeing-737-employees-messages.html>
- Gelles, D., & Kitroeff, N. (2019, October 30). Documents Show Safety Concerns at Boeing Before Deadly Crashes. *The New York Times*. Retrieved from <https://www.nytimes.com/2019/10/30/business/boeing-muilenburg-testimony-congress.html>
- Gelles, D., Kitroeff, N., Nicas, J., & Ruiz, R. R. (2019, March 23). Boeing Was 'Go, Go, Go' to Beat Airbus with the 737 Max. *The New York Times*. Retrieved from <https://www.nytimes.com/2019/03/23/business/boeing-737-max-crash.html>
- Gilbert, M. (1989). *On Social Facts*. London and New York: Routledge.
- Gilbert, M. (2013). *Joint Commitment: How We Make the Social World*. Oxford: Oxford University Press.
- Gowri, A. (2007). On Corporate Virtue. *Journal of Business Ethics*, 70(4), 391–400. doi:10.1007/s10551-006-9117-2
- Greco, J. (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press.
- House Committee on Transportation and Infrastructure. (2020). *The Boeing 737 MAX Aircraft: Costs, Consequences, and Lessons from Its Design, Development, and Certification—Preliminary Investigative Findings*. Retrieved from https://www.govinfo.gov/app/details/GOVPUB-Y4_T68_2-fb0f3812fefe3515ebcf3f4170fce64b
- Jones, T. (2007). Numerous Ways to Be an Open-Minded Organization: A Reply to Lahroodi. *Social Epistemology*, 21(4), 439–448. doi:10.1080/02691720701746565
- Kitroeff, N., & Gelles, D. (2019, April 20). Claims of Shoddy Production Draw Scrutiny to a Second Boeing Jet. *The New York Times*. Retrieved from <https://>

- www.nytimes.com/2019/04/20/business/boeing-dreamliner-production-problems.html
- Komite Nasional Keselamatan Transportasi. (2019). *Final Aircraft Accident Investigation Report* (Report No. KNKT.18.10.35.04). Retrieved from http://knkt.dephub.go.id/knkt/ntsc_aviation/baru/2018%20-%200035%20-%20PK-LQP%20Final%20Report.pdf
- Konzelmann Ziv, A. (2012). Institutional Virtue: How Consensus Matters. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 161(1), 87–96. doi:10.1007/s11098-012-9933-4
- Lackey, J. (2016). What Is Justified Group Belief? *The Philosophical Review*, 125(3), 341–396. doi:10.1215/00318108-3516946
- Lahroodi, R. (2007). Collective Epistemic Virtues. *Social Epistemology*, 21(3), 281–297. doi:10.1080/02691720701674122
- List, C., & Pettit, P. (2011). *Group Agency: The Possibility, Design, and Status of Corporate Agents*. New York: Oxford University Press.
- Moore, G. (2005). Corporate Character: Modern Virtue Ethics and the Virtuous Corporation. *Business Ethics Quarterly*, 15(4), 659–685. doi:10.5840/beq200515446
- Moore, G. (2015). Corporate Character, Corporate Virtues. *Business Ethics: A European Review*, 24(S2), S99–S114. doi:10.1111/beer.12100
- Odell, M. (2011, July 21). Boeing Bows to Pressure on 737 Upgrade. *The Financial Times*. Retrieved from <https://www.ft.com/content/cd0c2dabc-b2d0-11e0-bc28-00144feabdc0>
- Quinton, A. (1975). The Presidential Address: Social Objects. *Proceedings of the Aristotelian Society*, 76, 1–viii. doi:10.1093/aristotelian/76.1.1
- Sandin, P. (2007). Collective Military Virtues. *Journal of Military Ethics*, 6(4), 303–314. doi:10.1080/15027570701755505
- Sosa, E. (2007). *A Virtue Epistemology*. Oxford: Oxford University Press.
- Thomas, G. (2006). What Comes Next? *Air Transport World*, 43(5), 51–52.
- Weitzman, H. (2011, April 27). Boeing Dismisses Airbus Threat. *The Financial Times*. Retrieved from <https://www.ft.com/content/765db61e-70c8-11e0-9b1d-00144feabdc0>
- Zagzebski, L. T. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.

13b Commentary from Steven Bland

Getting More Out of Our Epistemic Vices

Summativism is the view that epistemic groups instantiate virtues and vices when a critical mass of their members do so. If epistemic virtues and vices scale up in this way, then collective virtues can be cultivated straightforwardly by cultivating those same virtues in constituent members. Barend de Rooij and Boudewijn de Bruin rightly reject summativism in favor of the view that groups instantiate epistemic virtues and vices when they function as epistemically virtuous or vicious agents. They argue that the functioning of an epistemic group depends not only on the behavioral tendencies of its members, but on the *organizational structures* that influence how its members *interact*. This is an important insight, but, as De Rooij and De Bruin recognize, the multiple realizability of virtues (and vices) presents new challenges to the project of promoting collective epistemic virtues. The three strategies they propose—virtue-to-function matching, organizational support for virtue, and organizational remedies against vice—strike me as plausible candidates. And they do an admirable job of showing how Boeing’s failure in all three respects contributed to their 737 Max disaster.

My aim in this commentary is not to undermine the strategies that De Rooij and De Bruin propose, but to point to a possible blind spot in their strategic approach more generally. Their approach, it seems, is to optimize the epistemic functioning of groups by maximizing and leveraging the virtues of their memberships. Virtuous organizations should design and implement policies, norms, and incentives that promote the epistemic virtues of their constituent members and mitigate their epistemic vices. On this view, the virtues and vices of a collective’s membership don’t scale up, but they do trickle up to the group level under the right organizational circumstances. We might call this a *virtue harnessing* approach. What it misses, in my view, is the potential that groups have to harness the epistemic *vices* of their members. This approach depends on what Smart calls our *Mandevillian intelligence*, which consists of “Cognitive and epistemic properties that are typically seen as shortcomings, limitations or biases at the individual level [that] can, on occasion, play

a positive functional role in supporting the emergence of intelligent behavior at the collective level” (Smart 2018, p. 4171). There are epistemic vices that neither scale up nor trickle up from individuals to groups, but instead function completely differently when manifested in solitary and group settings. Let’s review a few cases.

De Rooij and De Bruin discuss Lahroodi’s (2007) example of a narrow-minded group made up of open-minded members, but they don’t discuss the opposite situation of open-minded groups with closed-minded members. This is not only a live possibility, but an empirical reality. Most of us are closed-minded to one extent or another, which explains our susceptibility to a host of cognitive biases, including the pervasive confirmation bias. This disposition has epistemically deleterious effects on our reasoning in solitary settings: it can prevent us from properly justifying true beliefs and duly correcting false beliefs. On the other hand, Mercier and Sperber (2011) argue that its manifestation in dialogical conditions leads to a beneficial division of cognitive labor: every view under discussion gets tested by those who are most apt to find its faults (i.e., those who disagree with it), and defended by those who are most apt to find evidence in its favor (i.e., those who agree with it). There’s no one better to check our closed-mindedness than other closed-minded interlocutors. This is why, Mercier and Sperber claim, many of the biases that individuals exhibit when reasoning in isolation get drastically diminished, or disappear altogether, when biased minds reason together. Indeed, they claim that our biased minds evolved precisely because they led to our species’ considerable Mandevillian intelligence.

Collective deliberation is not an epistemic silver bullet, however. It is capable of opening our minds only when there is considerable diversity in the views under discussion. Deliberation within doxastically homogenous groups tends only to intensify our biases, leading to polarization, overconfidence, and belief perseverance. Consequently, De Rooij and De Bruin are quite right to emphasize that “*Organizational opposition or dissent is an essential macro-level remedy against various epistemic vices.*” But dissent can be difficult to *maintain* within an organization. Convergence toward a consensus is often a good thing, but not when it happens before every view has a fair and thorough hearing. And social cascades are constantly threatening to prematurely collapse viewpoint diversity. One possible safeguard against this form of groupthink is the inclusion of *overconfident* group members who privilege their own views over the contributions of others (Bernardo and Welch 2001; Zöllman 2010). Of course, these individuals should not be epistemically unreachable, but just confident enough to break up cascades that would foreclose discussion on what should be live options. In short, overconfidence and belief perseverance can be collective epistemic assets when they *precede* deliberation, rather than resulting from it.

I have little doubt that De Rooij and De Bruin's virtue harnessing strategies are conducive to the epistemically virtuous functioning of collective agents. My plea is only to expand the space of possibilities to include some vice harnessing strategies that facilitate a group's Mandevillian intelligence. One such strategy might be to promote and support adversarial deliberation among closed-minded, sometimes overconfident, group members.

References

- Bernardo, A.E. and Welch, E. (2001). On the evolution of overconfidence and entrepreneurs. *Journal of Economic & Management Strategy*, 10(3): 301–330.
- Lahroodi, R. (2007). Collective epistemic virtues. *Social Epistemology*, 21(3): 281–297.
- Mercier, H. and Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2): 57–74.
- Smart, P.R. (2018). Mandevillian intelligence. *Synthese*, 195: 4169–4200.
- Zollman, K. (2010). The epistemic benefits of transient diversity. *Erkenntnis*, 72: 17–35.

13c Commentary from Neil Levy

Barend de Rooij and Boudewijn de Bruin aim to extend virtue (and vice) epistemology into an analysis of the epistemic dispositions possessed by groups. In their chapter, they focus on the epistemic vices that (they claim) contributed to two deadly aircraft crashes. Multiple failings by Boeing, aimed at speeding certification for the aircraft involved in both crashes and at keeping the cost of the planes low, resulted in corners being cut with disastrous consequences. De Rooij and de Bruin argue that these failings were at least partly epistemic: they included failures to share information with the regulators and arrogance about beliefs held. De Rooij and de Bruin hold, in addition, that these epistemic failures should be seen as manifestations of genuinely collective epistemic vices.

I am agnostic on whether groups like corporations can helpfully be conceptualized as possessing epistemic virtues and vices. There's no barrier in principle to such ascription: I share the view that such states should be analyzed as functional states, and groups can certainly possess such states in principle. The issue for me is empirical: are such groups really constituted in the right way for such conditions to be realized? While there may be more than one way to meet these conditions, it's likely that possession of these states will require that the group be constituted such that it has sensors of some kind by which information is received, internal processing mechanisms, and further mechanisms for outputting the transformed information, and that it possesses these mechanisms *qua* group. At minimum, that seems to require a high degree of integration at the processing stage. I doubt many groups satisfy these conditions, but some might. Perhaps Boeing (or, more plausibly, some decision-making group within Boeing) satisfies these conditions.

I am, however, deeply resistant to the idea that we should expect corporations to manifest the virtues in the way de Rooij and de Bruin call for. While they reject summativism, the view that talk of group virtues can best be understood as mere shorthand for the virtues of (key) individuals, nevertheless they do see the genuinely collective virtues they ascribe as in some manner depending on individual virtues. While they deny that an epistemically open-minded organization must be composed of epistemically open-minded individuals, they nevertheless believe that

group-level virtue depends on individual-level virtue. For them, epistemically virtuous organizations must have some way of ensuring that “group members exhibit the virtues demanded by their roles within the organization” (emphasis added). This is a claim they repeat several times, in slightly different terms: it must be the case that “the members of that organization have the epistemic virtues required by their roles and responsibilities within the organization.” Boeing, they say, failed to “harness the virtues” of its engineers; it did not sufficiently resemble one of those organizations that provide “supportive environments to enable epistemically virtuous individuals.” Nor did its CEO manifest the epistemic virtues. I am skeptical that we should ask or expect those within an organization to manifest the virtues.

Perhaps in an ideal world, we might make such a demand of employees. In the actual world—the world on which, I take it, they hope to have an impact—such a demand should be rejected.

We already expect individuals to subject themselves to the often arbitrary and dictatorial rule of what Anderson (2017) calls the “private government” of employers. More and more, employees are policed by their employers outside the workplace as well as within. Within the workplace, their time, their dress, and their behavior are tightly controlled. We should not also ask them to behave virtuously. That would represent a further creep of the corporation into our mental lives that should be resisted. As employees, we have duties, moral and legal, to our employers and to one another. We should resist anything more demanding than the requirement that we act in accordance with these duties.

We shouldn’t, for example, expect employees to be open-minded in appraising proposals, or intellectually generous. Open-mindedness and generosity may be great things (or they may not; see Levy 2002; Levy and Alfano 2020). But they depend on freedom of thought that is at odds with the hierarchical nature of employment and on a commitment we should not ask of workers. In an ideal world, a world in which labor is not alienated, we might make such a demand of one another, but that is not *this* world.

CEOs are a different matter. They are controllers, not the controlled, and I have no qualms in asking them to manifest the virtues. Nevertheless, I am deeply skeptical we’ll make a dent in the kinds of problems that motivate de Rooij and de Bruin that way. There is, of course, wide variation in how ethical CEOs are, but we should recognize few of them are likely ever to be exemplars of virtue. Typically, neither members of the pool of potential CEOs nor the boards of organizations that appoint them have any genuine interest in even minimal decency, and looking to them for better behavior is always going to be a futile exercise.

Employees should not be expected to turn over any more of their mental lives to corporations than they already do and CEOs will not become virtuous any time soon. We do far better to ensure genuinely effective

regulation of how corporations behave than to attempt to inculcate the virtues. We need proper protection for whistleblowers, regulators who are genuinely independent of the industry they regulate, mechanisms to shut the revolving door between industry, government, and oversight. We also need effective sanctions for those that violate the regulations. These kinds of measures are far more likely to be effective at minimizing the harms that corporations do than any attempt at inculcating the virtues in their members.

References

- Anderson, E. 2017. *Private Government*. Princeton, NJ: Princeton University Press.
- Levy, N. 2002. Against Philanthropy, Individual and Corporate. *Business & Professional Ethics Journal* 21: 95–108.
- Levy, N. & Alfano, M. 2020. Knowledge from Vice: Deeply Social Epistemology. *Mind* 129: 887–915.

13d Barend de Rooij and Boudewijn de Bruin's Response to Commentaries

When complex organizations fail, it can be challenging to determine which of their component parts fall short. Take the Dutch Tax and Customs Administration, a crucial player in the “child benefits scandal” that forced the Dutch government to resign in January of 2021, and which made headlines even in the *New York Times* (Erdbrink, 2021). Tasked with overseeing the distribution of childcare benefits, the organization is also responsible for discovering cases of fraud among beneficiaries. It is at this task that the organization failed spectacularly, wrongly accusing thousands of vulnerable parents of making fraudulent claims and requiring them to pay hefty fines.

After a parliamentary inquiry into the child benefits scandal vigorously condemned the “unprecedented injustice” done to the families involved (Parliamentary Committee of Inquiry, 2021), Prime Minister Mark Rutte was quick to apologize for “mistakes ... made on all levels” (Erdbrink, 2021). And indeed, the causes of the tax authority’s failing are manifold. Politicians gave the organization a strong mandate to uncover fraud, contributing to its zealotry. Institutional biases hindered the organization’s working procedures. Case administrators, legal professionals, and other government employees were swept up in the organization’s performance-oriented culture, wrongly deciding numerous fraud investigations.

Given this complexity, it would be too easy to blame the scandal on a single party. The government has rightly accepted accountability, but the actions of individual employees have also come under scrutiny. Government workers, critics have argued, must “sharpen their moral compass” (van den Berg et al., 2021). We venture to say that cultivating the *epistemic* compass should get priority.

What would we recommend more concretely? In our chapter, we distinguish three independent sources of corporate epistemic virtue. First, individuals should have the epistemic virtues that match the roles they play. Second, organizations should support the exercise of epistemic virtue. And third, organizations should remedy epistemic vice.

According to Neil Levy, we should not ask employees to manifest epistemic virtue, except “[p]erhaps in an ideal world.” We agree with

Levy that where employee–employer relationships suffer the repressive consequences of hardnosed hierarchical management views, exhorting employees to virtue may be as cynical as it is useless. However, our evaluation of working conditions generally leads us to a more optimistic appraisal of the benefits of stimulating corporate virtue the way we envisage it. In fact, we think that corporate epistemic virtue diffuses the sharper edges of hierarchy; that is, corporate epistemic virtue guards against what Levy calls the “creep of the corporation in our mental lives.”

To appreciate this, it is important to recall that our view of corporate epistemic virtue entails not only a recommendation to match virtues to functions, but also a recommendation that organizations seeking corporate epistemic virtue put in place *support* for epistemic virtue and *remedies* against epistemic vice. While Levy appears to suggest otherwise, the kind of measures he mentions are exact examples in which support and remedies are absent: insufficient whistleblower protection, lack of regulator independence, the revolving door—they all stand in the way of exercising open-mindedness, love of knowledge, epistemic justice, and a host of other epistemic virtues. Better regulation, then, stimulates corporate epistemic virtue by facilitating the exercise thereof (De Bruin, 2021).

But should we sometimes advise epistemic vice? It seems we should. One of the Members of Parliament who brought the problems at the Tax Office into the spotlight has been portrayed in the media as narrow-mindedly obsessed with the case. He was often characterized as a nuisance whose overly critical questions threatened to derail effective government (Klaassen, 2021). Nevertheless, his narrow-minded focus was essential in raising awareness about the scandal among other Members of Parliament, paving the way for parliament-led corrective action.

We therefore agree wholeheartedly with Steven Bland’s suggestion that the epistemic *vices* of group members can be conducive to collective virtue as well. When particular epistemic vices contribute to collectively beneficial outcomes, it is a good idea to try and use them to their maximum potential. In earlier work, one of us used research on CEO overconfidence to illustrate this point. We take the opportunity to summarize it briefly. It is well known from the economics and business literatures that firms tend to hire overconfident CEOs. *Prima facie* that does not make sense from an orthodox expected utility maximizing point of view, and Hirshleifer and colleagues describe this observation, a little hyperbolically, as a the “biggest puzzle raised by existing research on managerial beliefs and corporate policy” (2012, p. 1459). Their original solution? Overconfidence in CEOs benefits a company, as they are greater innovators. Balanced by a more reticent and cautious CFO and other directors, the “visionary” CEO leads the firm to great heights.

But how far should we go? Ultimately what mix of epistemic virtue and vice, and what forms of support for virtue and remedy against vice

will be needed is an empirical question, which De Bruin (2015) has started answering on the basis of recent work in behavioral finance. Overconfident CEOs may be good innovators, but there is equally high-caliber evidence that CEO overconfidence may have a negative impact on shareholder value and corporate investment decisions (Malmendier and Tate, 2005). When an organization aims to harness the vices of its members, then, it must be careful to ensure that this strategy promotes its intended target.

Bibliography

- De Bruin, B. (2015). *Ethics and the Global Financial Crisis: Why Incompetence Is Worse than Greed*. Cambridge: Cambridge University Press.
- De Bruin, B. (2020). Epistemic Corporate Culture: Knowledge, Common Knowledge, and Professional Oaths. *Seattle Law Review*, 43(2), 807–839. <https://digitalcommons.law.seattleu.edu/cgi/viewcontent.cgi?article=2654&context=sulr>
- Erdbrink, T. (2021, January 15). Government in Netherlands Resigns After Benefits Scandal. *The New York Times*. Retrieved from <https://www.nytimes.com/2021/01/15/world/europe/dutch-government-resignation-rutte-netherlands.html>
- Hirshleifer, D., Low, A., & Teoh, S. H. (2012). Are Overconfident CEOs Better Innovators? *The Journal of Finance*, 67(4), 1457–1498. <https://doi.org/10.1111/j.1540-6261.2012.01753.x>
- Klaassen, N. (2021, April 26). Notulen openbaar: ministers moesten ‘activistische’ en ‘tegenwerkende’ Kamerleden terugfluiten. *Het Parool*. Retrieved from <https://www.parool.nl/nederland/notulen-openbaar-ministers-moesten-activistische-en-tegenwerkende-kamerleden-terugfluiten~be6e63af/>
- Malmendier, U., & Tate, G. (2005). CEO Overconfidence and Corporate Investment. *The Journal of Finance*, 60(6), 2661–2700. <https://doi.org/10.1111/j.1540-6261.2005.00813.x>
- Parliamentary Committee of Inquiry into Childcare Benefit. (2021). Ongekend Onrecht: Verslag—Parlementaire ondervragingscommissie Kinderopvangtoeslag. Retrieved from https://www.tweedekamer.nl/sites/default/files/atoms/files/20201217_eindverslag_parlementaire_ondervragingscommissie_kinderopvangtoeslag.pdf
- van den Berg, I., van de Bunt, J., Kuiper, G., van de Luijngaarden, E., Mein, A. (2021, May 26). Juristen bij uitvoeringsorganisaties moeten hun moreel kompas scherper stellen. *Trouw*. Retrieved from <https://www.trouw.nl/opinie/juristen-bij-uitvoeringsorganisaties-moeten-hun-moreel-kompas-scherper-stellen~bedfce965/>

14 The Social Virtue of Questioning

A Genealogical Account

Lani Watson

Questioning is an integral part of our lives. It features in our social interactions in myriad ways, allowing us to connect and coordinate with each other in public and in private. Questioning arises across cultures, histories and social contexts, binding us to common goals, establishing common ground, allowing us to challenge and provoke, and serving as a vital tool in the search for new and useful information. Of course, it plays an important role in philosophy too, a discipline that famously ‘begins in wonder’ (Plato, *Theaetetus* 155d and Aristotle, *Metaphysics*, Book 1A, 982b12). After all, philosophers wouldn’t get very far if they failed to move beyond wonder and start asking questions. In short, questioning is familiar, indispensable and ubiquitous.

Despite its significance for both philosophy and everyday life, limited philosophical attention has been paid to questioning. Only in the latter half of the twentieth century has there been any sustained philosophical inquiry into the nature of questions or questioning and much of this work has focused on the formal logical or linguistic analysis of questions (Prior and Prior 1955; Åqvist 1965; Belnap and Steel 1977; Karttunen 1977; Hintikka 1983; Ginzburg 1996; Higginbotham 1996; Groenendijk 1999; Aloni 2005; Jaworski 2009; Ciardelli 2010). This has undeniably provided valuable insights but the logical and linguistic analysis of questions captures only one aspect of an expansive philosophical landscape in which questioning operates as an *epistemic practice*.

In this chapter, I offer an account of questioning as an epistemic practice. I approach this by providing a genealogical account of questioning, inspired by the genealogical account of knowledge presented by Craig (1999). The genealogical approach serves, in the first instance, as a basis for establishing and evaluating the descriptive claim that questioning is an epistemic practice. In the second instance, it provides the impetus for normative analysis of questioning. In both cases, the genealogical approach helps to elucidate the social role and function of questioning. As the title indicates, this leads to an account of the social virtue of questioning. I conclude the chapter with a brief discussion of the prospects

for a contemporary epistemology of questioning in which I seek to highlight, in particular, its significance at the intersection of social and virtue epistemology.

1 What is a practice

Questioning is an epistemic practice. In order to understand what this means, we must first interrogate the notion of practices, in general, and then ask what makes a practice epistemic. The notion of practices, in general, has attracted attention in contemporary academic discourse across a range of disciplines including philosophy, sociology, political science, psychology and anthropology (this has been labelled the ‘Practice Turn’ (Schatzki 2001)). Within this diverse literature, the notion of practice has been given various and sometimes conflicting characterisations. Commenting on the varied interpretations found in the literature, Rouse (2006) observes, “[P]ractices range from ephemeral doings to stable long-term patterns of activity” (p. 499). Nonetheless, a central feature of most, if not all theories of practice is the identification of practices as activity based. As Schatzki (2001) writes, “[M]ost thinkers who theorize practices conceive of them, minimally, as arrays of activity” (p. 11). Practices are essentially constituted by activities and it is primarily on this basis that they are distinguished from theories.

Take an everyday example, say, the practice of holding a door open for the person entering a building behind you. This practice is primarily constituted by the action of holding open the door. Without this action, no amount of theorising about the practice will bring it into existence as a practice. Moreover, the notion of a practice extends beyond the performance of an individual action. Practices incorporate the repetition or reiteration of an action or set of actions over a sustained period. If a door is held open only once and the action is never repeated, no practice of holding doors open can be said to exist. Determining exactly how much repetition or reiteration is required for a practice to emerge is a topic for a more detailed study. It will suffice here to note that practices are more than individual actions.

Similarly, a practice requires more than the unreflectively coordinated actions of groups of individuals; individuals must be acting together or cooperating in some meaningful sense. If doors were held open at random without the basic aim of easing passage into buildings, then the holding open of doors could not be called a practice. As such, while a practice may incorporate or be akin to a custom or habit, practice theorists typically extend the notion beyond custom or habit arguing, in particular, that practices are defined by the common goals towards which the actions involved are directed (Barnes 2001; Turner 2001). A set of coordinated actions with no common goal, such as the aimless holding

open of doors, will not amount to a practice (although it may amount to a custom or habit). Practices are sets of activities that arise among groups of individuals with common goals.

Such groups typically constitute a society or community. Practices are, therefore, importantly social. The practice of holding doors open is a prime example; it serves as a means of structuring and coordinating interactions between members of a community. The social context thus plays an important role as a basis for practices. MacIntyre (1981) emphasises this in his work on moral and political practices: “[B]y a ‘practice’ I am going to mean any coherent and complex form of socially established cooperative human activity” (p. 175). Practices serve as a means of organising and structuring social spaces and interactions. Without this social context, the notion of practice makes little sense. If you have ever visited a country where there is no practice of holding doors open and have nonetheless attempted to engage in it, you will have experienced the significance of the societal uptake of practices first-hand.

This social basis does not preclude practices being enacted in private. Indeed, many practices are performed by individuals operating outside of an explicitly social context, including spiritual practices, such as meditation or prayer. These practices nonetheless arise out of and in response to a social context that supplies the parameters for their performance in private. While ostensibly taking place in private, individuals engaging in a practice such as meditation or prayer are still operating under, and are therefore constrained by, the social context in which the practice originally emerged. Several key ideas from the practice theory literature can thus be brought together to arrive at a broadly uncontentious account of practices: practice is a *socially established set of activities directed towards common goals*.

2 What makes a practice epistemic

Questioning is not merely a practice; it is an epistemic practice. In order to determine what makes a practice epistemic (and so what makes questioning an epistemic practice), it will be useful to briefly characterise ‘the epistemic’ in general terms. This is often done in contemporary epistemology by providing a characterisation in terms of a collection of states or goods, including, at least, true belief, justification, information, knowledge, and understanding. This approach picks out an intuitively coherent set of ‘epistemic goods’. As Baehr (2011) comments, “[W]hile more could be said to demarcate epistemic ends from other kinds of ends, the basic distinction should be intuitive enough” (p. 209). I will adopt this intuitive approach and treat ‘epistemic’ as a catch-all term for a collection of states or goods, including, at least, true belief, justification, information, knowledge and understanding.

We can feed this characterisation of the epistemic into the account of practices already given. A practice is a socially established set of activities directed towards common goals. The goal-directedness of practices is an important feature. In particular, practices can be demarcated by reference to the goals at which they aim. Spiritual practices, for example, aim at distinctively spiritual goals, such as enlightenment or communication with the divine. In much the same way, epistemic practices have distinctively epistemic goals; they aim at epistemic goods, such as true belief, justification, information, knowledge and understanding. It is directedness towards these epistemic goals that makes a practice an epistemic practice.

This strategy of demarcating practices in terms of goals has been adopted throughout the literature. Alston (1989), for example, comments, “[A] doxastic practice can be thought of as a system or constellation of dispositions or habits...each of which yields a belief as output” (p. 5). Here, a doxastic practice is characterised in terms of the goal of belief. Similarly, Roberts and Wood (2007) state, “[I]ntellectual practices aim intrinsically at such goods as understanding... acquaintance, and confirmation of beliefs” as well as “the justification and warrant of beliefs” (p. 117). Once again, intellectual practices are characterised in terms of the intellectual goals at which they aim.¹ Likewise, epistemic practices aim at epistemic goals and are distinguished from other practices on this basis. An epistemic practice is *a socially established set of activities directed towards common epistemic goals*.

3 What makes questioning an epistemic practice

We now have an account of practices and an account of what makes any particular practice an epistemic practice: a practice is an epistemic practice just in case it constitutes a socially established set of activities directed towards common epistemic goals. We can thus interrogate the claim that questioning is an epistemic practice according to three definitive criteria: (1) Is questioning socially established, (2) Is questioning activity based, (3) Is questioning directed towards common epistemic goals. The answer to each of these questions is yes. In order to provide support for this, I will sketch out a genealogical account of questioning. A full-bodied genealogical account would require significantly more space than is afforded here but the following sketch will nonetheless helpfully elucidate both the social nature and function of questioning. In turn, this will lead to an account of the social virtue of questioning and its significance at the intersection of social and virtue epistemology.

Before proceeding, a terminological point will be helpful. I will characterise the social domain in which questioning takes place as an *epistemic community*. An epistemic community consists of a group of individuals that produces, shares and consumes epistemic goods, such as

information, knowledge and understanding. Epistemic communities are a central feature of human life and progress. As observed by Goldman (1999), “[A] hallmark of human culture...is to enhance the social fund of knowledge by sharing discovered facts with one another” (p. 103). Throughout history, epistemic communities have arisen and flourished on increasingly grander scales. The epistemic communities of ancient Athens and Rome produced and nurtured intellectual advancement through the exchange of ideas and innovations by means of oratory and written records. The fifteenth-century epistemic community of Europe expanded dramatically as a result of the invention of the printing press and the subsequent Printing Revolution. In its contemporary manifestation, the epistemic community is, in essence, a global one with the advent of the World Wide Web taking centre stage in the increasingly rapid rate of information exchange.

Notably, in all these cases, the epistemic community does not consist merely in a network of epistemic goods and epistemic agents, however large: epistemic communities are dynamic. In order to constitute an epistemic community, epistemic goods must be *exchanged* between members. This exchange of epistemic goods – beliefs, information, knowledge, etc. – forms, shapes and sustains our epistemic communities. We “enhance the social fund of knowledge”, as Goldman puts it, “by *sharing* discovered facts with one another” (p. 103, emphasis added). Crucially, within an epistemic community, the exchange of epistemic goods is frequently facilitated by the asking and answering of questions. It is this important role, then, that will be elucidated by a genealogical account of questioning.

4 The genealogy of questioning

A genealogical account of questioning examines the role or function of questioning in our epistemic communities. In essence, it is an account of why people ask questions. Answering this is key to identifying questioning as an epistemic practice. Examining the nature of a thing in terms of its function, moreover, is an approach familiar across both philosophical and scientific disciplines. Biologists, for example, typically identify the heart in terms of its function of pumping blood around the body. Functionalists, within the philosophy of mind, argue for the same approach to the mind.

Within epistemology, this function-based approach has been advocated by Craig in his influential book, *Knowledge and the State of Nature* (1999). Craig proposed that an investigation into the function of the concept knowledge would yield important insights into the nature of knowledge itself. Thus, he argues:

There seems to be no known language in which sentences using ‘know’ do not find a comfortable and colloquial equivalent. The

implication is that it answers to some very general needs of human life and thought, and it would surely be interesting to know which and how.

(Craig 1999, 2)

In essence, Craig's genealogical account examines why the concept knowledge has emerged in human societies and what purpose it serves. An equivalent approach, in the case of questioning, looks promising given, as already noted, the ubiquity and social nature of the practice. Questioning features in our social interactions in myriad ways. It spans cultural and linguistic boundaries and operates within a wide variety of diverse and distinct social contexts. Questioning has, moreover, been a feature of human social interactions since the beginning of recorded history and, one may well imagine, extending into prehistory. A genealogical account of questioning thus offers a promising route to understanding the social nature and function of questioning as an epistemic practice. To arrive at this account, it will be helpful to look at the genealogical account of knowledge in more detail.

Craig (1999) begins his genealogical account of knowledge with a thought experiment. He imagines a society in which the concept 'knowledge' does not yet exist. He then asks why such a society would develop the concept and what function it would have. In answer to these questions, Craig focuses on the role that knowledge ascriptions play in identifying individuals who possess information within a community. He argues that a key concern for members of any community is the possession of true beliefs. Consequently, communities require sources of information by which they can form true beliefs. Firstly, Craig acknowledges the sources of information that people have at their immediate disposal, including their perceptual and mental faculties. Secondly, he highlights the advantages of being able to consult and utilise the faculties of others. If I am unable to see what's around the next corner, I can rely on the people walking in front of me to determine whether or not there are dangers ahead. As Craig puts it, "the tiger that Fred can see and I can't may be after me and not Fred" (1999, 11).

With the significance of these secondary information sources in mind, Craig proposes that the concept of knowledge would emerge in the society of the thought experiment as a means of identifying useful sources of information. Thus, he argues, "the concept of knowledge is used to flag approved sources of information" (1999, 11). Picking Fred out as someone who *knows* a lot about the local tigers is a useful way of identifying him as a good source of information about the local tigers and, therefore, as a good candidate for leading the trek through the jungle. According to Craig's genealogical account, knowledge ascriptions are a means of identifying good informants within an epistemic community.²

As noted earlier, epistemic communities are dynamic. They do not merely consist in networks of good (or bad) informants but require the exchange of epistemic goods between informants in order to form, persist and grow. It is by shifting perspectives on the Craigian account of knowledge, then, that we can develop a genealogical account of questioning. Indeed, both accounts begin with the very same thought experiment. Imagine a society in which the practice of questioning did not exist, then ask why the practice would emerge in such a society and what purpose it would serve. This small but significant change in perspective forces us to appreciate the dynamism of the epistemic community in Craig's original account. There is limited value in using knowledge ascriptions to flag good informants, if one cannot go on to access and ultimately benefit from the information that they have. Questioning is not the only method for accessing that information but, as I will argue, it is a powerful and pervasive one.

As with Craig's (1999) original account, the genealogy of questioning begins with the grounding notion that a key concern for members of any community is the possession of true beliefs. One highly effective way of coming to possess true beliefs is by seeking them out. This requires seeking out the information on which one's true beliefs will be based. As such, information seeking is an essential activity in the successful functioning of any community. As Goldman (1999) observes, "[I]nformation seeking is a pervasive activity of human life" (p. 3). We seek out information on a regular basis and use it in all manner of ways to determine how to act. It is difficult, in fact, to see what our lives would look like if we did not. It is no good, for example, simply knowing that Fred is a great source of information about the local tigers, if one is concerned that there may be a tiger around the next corner. The ability to access the information Fred has at his disposal is vital for deciding what to do and, at least in this case, improving one's chances of survival. A community in which information seeking did not take place would plausibly not last long.

The process of information seeking, moreover, requires some sort of mechanism. Craig acknowledges the role that perceptual and mental faculties play as sources of information. In addition, however, they play a role as *information-seeking* mechanisms. Furthermore, as noted, Craig emphasises the advantages of being able to consult and utilise the faculties of others. Questioning enters the spotlight by performing precisely this function. Questioning allows us to consult and utilise the faculties of others and so to seek out the information that we need or want in order to form true beliefs and decide how to act. How am I to discern whether it is safe to walk through the jungle? I *ask* Fred if he can see any tigers up ahead or, perhaps more wisely, if he thinks there is likely to be one in the area. Of course, it's possible that Fred will provide me with this information without me having to ask for it. Nonetheless, my ability

to ask for the precise information that I need, at the precise moment that I need it, significantly improves my chances of making an informed (and potentially life-saving) decision.

Naturally, not all of our information-seeking activities are as vital and urgent as those involving tigers and jungles. Nonetheless, much of the information-seeking that we do on a daily basis is driven by minor but relevant needs: what time is the meeting; when is the next bus due; how long will it take to get there. Questioning provides an effective and efficient means of accessing the information that we need at the time that we need it, often by reaching out to others who already have the information at their disposal or can more easily acquire it. A society that did not engage in questioning would plausibly be at a significant disadvantage, both with respect to the basic continued survival and everyday functioning of its members. We thus have an answer to the Craigian-style question of why the practice of questioning would emerge in a society in which it did not yet exist. Simply put, information seeking is an essential activity in the successful functioning of any society and questioning is a powerful and pervasive information-seeking mechanism.

This almost completes our sketch of the genealogy of questioning. A small refinement will be useful before proceeding to an account of the social virtue of questioning. Specifically, note the distinction between *seeking* and *eliciting* information. Seeking information refers to the act or process of searching for it, while eliciting information captures the sense in which information, once sought, is also acquired. If one seeks information, one may coherently fail to acquire it. If one elicits information, then one in fact acquires it. When we ask questions, we don't simply aim to search for information, we aim to search for *and* acquire it. That is not to say that we always succeed in acquiring information when we ask questions but merely that we are trying to do so. As such, the role or function of questioning in our epistemic communities is best characterised as information elicitation. From hereon I will talk of information elicitation, rather than information seeking.

Interestingly, this subtle distinction helps to bring an important objection to the fore. One might be concerned that the genealogical account of questioning just sketched is too narrow. In particular, in this account, questioning is understood exclusively in terms of the goal of eliciting information. Yet, we use questions to achieve many other goals, above and beyond searching for and acquiring information. Indeed, one may imagine any number of scenarios in which questioning plausibly takes place with some other, perhaps explicitly non-epistemic goal in mind. We may, for example, question in order to demonstrate care or concern for another, or to provoke a response such as surprise or embarrassment. Thus, one could argue that the goal of questioning is not always or necessarily to elicit information. This, in turn, puts pressure on the claim that questioning is an epistemic practice (because epistemic practices have epistemic goals).

There is much that can be said in response to this objection and, unfortunately, not enough space to address the concern in detail here. However, the genealogical approach provides us with an indication of the relevant response. Remember that this approach asks why questioning would emerge in a society in which it didn't already exist and what purpose it would serve. We can ask this of any number of things and those who study and write about our prehistoric past often do: sociologists, anthropologists, archaeologists and so on. In essence, these questions constitute two ways of asking the same basic thing: what does it do? Whether we are asking this of a concept like knowledge, a practice like questioning, or an archaeological find like a stone-age tool, the question focuses us on the purpose or function of the thing in question. This focus aligns with the genealogical approach, broadly speaking.

With this approach in mind, it is useful to conceive of questioning as a tool (perhaps even a stone-age one!). Much like other tools, questioning is characterised in terms of what it does. A wrench is a tool for tightening and untightening nuts and bolts, a hammer is a tool for hammering in nails, and questioning is a tool for eliciting information. Crucially, as with all tools, questioning can be used for any number of other purposes. A wrench can be used to prop open a door, a hammer to smash through a glass pane, and so on. Likewise, questioning can be used to demonstrate care or concern for another, or to provoke a response such as surprise or embarrassment. We nevertheless characterise questioning, like all tools, in terms of its basic or primary function. A wrench is not a wrench in virtue of its ability to prop open doors, it is a wrench in virtue of its primary function for tightening and untightening nuts and bolts. Likewise, questioning is not questioning in virtue of the fact that it can be used to care for or embarrass someone. The fact that it can be used for these, and numerous other non-epistemic purposes, does not suffice to undermine the primary epistemic function of questioning: that of eliciting information.

The genealogical account provides a plausible story in support of this account of the primary epistemic function of questioning. Questioning would emerge in a society in which it did not already exist in order to provide a mechanism for the exchange of information within an epistemic community. Not as a means of demonstrating care or provoking embarrassment, even though it can be used to achieve these secondary goals. If these were in fact the primary goals of questioning, and the information eliciting function was absent, then we would be living in a world in which questioning was *entirely* rhetorical. Rhetorical questions are defined precisely by the absence of an information eliciting function (Watson, 2021).

In such a world, there may emerge some other mechanism for exchanging information within an epistemic community or perhaps we would rely solely on our individual perceptual and mental faculties. But this

would not be a world in which questioning proper (i.e., non-rhetorical questioning) would have emerged. Questioning is rightly characterised in terms of the epistemic goal of eliciting information. As noted, a deeper engagement with this issue is ultimately required in order to demonstrate the full force of the position outlined here. Nonetheless, I believe the genealogical account offers a compelling context for understanding and evaluating the practice of questioning in terms of the information eliciting goal.

We can now return to the claim that questioning is an epistemic practice. I noted three definitive criteria for determining this: (1) Is questioning socially established, (2) Is questioning activity-based and (3) Is questioning directed towards common epistemic goals. The genealogical account supports the answer yes, in each case. According to this account, questioning plays an important social role in our epistemic communities as a mechanism for information elicitation. As such, it emerges from and within our social world and serves as a means of structuring interactions between members of our epistemic communities, at the same time sustaining and expanding them. As with the practice of holding open doors, questioning serves to structure and organise interactions between individuals in a social setting; it is socially established.³

Similarly, as with the practice of holding open doors, questioning is activity-based. The activity that makes up questioning is, perhaps somewhat mundanely, the asking of questions. Such asking can take various forms and there is plenty more to be said about these (Watson, 2021), but for present purposes ‘the asking of questions’ will suffice for an uncontentious description of the activity that constitutes questioning. Without this activity, no practice of questioning could be said to exist. Again, as with the practice of holding open doors, the asking of questions has a common (and primary) goal. The common goal of questioning is that discussed and defended above: the goal of eliciting information. Given that information is an epistemic good, the goal of eliciting information is rightly characterised as an epistemic goal. Questioning is, therefore, an epistemic practice. It is *a socially established set of activities directed towards the common goal of eliciting information.*

5 The social virtue of questioning

I have argued that questioning is an epistemic practice. This is, in essence, a descriptive claim. Beyond this descriptive claim, the genealogical approach I have adopted naturally lends itself to a further examination of the practice of questioning in normative terms. Specifically, on the basis of this approach one can easily begin to appreciate the significant social and societal value of questioning and I will now seek to further elucidate this value. This is what I mean by giving an account of the social virtue of questioning. In fact, one could substitute virtue for value here. Both

of these words indicate the normative dimension that we are now turning to with respect to questioning in the social domain. Following this normative examination, in the final section, I will go on to emphasise the relevance of questioning as a topic of interest at the intersection of contemporary social and virtue epistemology.

We can begin the normative examination of questioning by returning to a point made at the start of the chapter. Much of the existing work on questions in philosophy has focused on the logical and linguistic analysis of questions, in either formal, syntactic, or semantic terms (Prior and Prior 1955; Åqvist 1965; Belnap and Steel 1977; Karttunen 1977; Hintikka 1983; Ginzburg 1996; Higginbotham 1996; Groenendijk 1999; Aloni 2005; Jaworski 2009; Ciardelli 2010). This is not, of course, the only approach to be found when searching through pockets of the literature in various sub-disciplines (e.g., a different lens is adopted in the philosophy of science where the focus is scientific inquiry or method, broadly speaking [Van Fraassen 1980; Hintikka 1981; Koura 1988]). But it is fair to say that this, broadly speaking, formal approach has been dominant in the philosophical study of questions, thus far (see the *Stanford Encyclopedia of Philosophy* entry on ‘Questions’ for an overview (and confirmation) of this (Cross and Roelofsen 2018)).

By viewing questioning as an epistemic practice, however, we add an important dimension to the formal study of questions. In fact, one might argue that the form or structure of a question is significantly dependent on and determined by the practice of questioning. Rather than existing independently of this practice, questions necessarily operate within, and are therefore also constrained by, the practice of questioning: they are the activity that constitutes the practice. As such, questions – as a form of language – are embedded within the practice of questioning, which is itself governed by a set of norms. This close relationship between language and practice was emphasised by Wittgenstein in the *Philosophical Investigations* (1953). Wittgenstein argued for the priority of practices in any rule-following system and applied this idea, particularly to language use. Thus, he maintained that formal linguistic conventions arise out of rather than determine linguistic practices. Plausibly, then, the social context in which the practice of questioning takes place is essential to understanding the form and structure of questions. In order to gain a rich understanding of questions, we should treat them within the broader context of the practice of questioning.

Moreover, by viewing questioning as an epistemic practice, we not only inform formal analyses of questions as a linguistic expression but can move beyond these into an examination of a ubiquitous and indispensable activity of everyday life. As I’ve suggested, this brings with it a rich normative dimension. Viewing questioning as a practice allows and, indeed, I think requires us to consider the social, political and educational roles and effects of questions and questioning. What does it mean

to be a questioner? How do questions help or hinder social, political, or scientific progress? Who gets to ask questions and who doesn't? Can we teach good questioning? These are important questions that should motivate attending to the normative aspects of questioning as an epistemic practice: the social virtue of questioning.

What, then, is the social virtue of questioning. As I have argued, questioning occupies a vital place in our epistemic communities. It plays a central role in our social epistemic interactions, providing a highly effective means of eliciting precisely the information that we need or want, at precisely the moment we need or want it. On this basis, the value of questioning from a social, or perhaps societal, perspective is not hard to see. Questioning allows us to both access and generate epistemic goods such as true belief, justification, information, knowledge and understanding – goods that we value within our epistemic communities. Equally, if not more importantly, questioning facilitates the exchange of these goods among community members – the *sharing* of information, knowledge and so on. As such, questioning serves to ease the passage of epistemic goods between members of an epistemic community. Much as the holding open of doors eases passage in and out of buildings.

Indeed, it is hard to see how the smooth and efficient exchange of epistemic goods would be possible in the absence of questioning. How would one go about signifying desire for a particular piece of information, or be able to identify it in others, without the use of a question? To be clear, my claim is not that these things would be impossible in the absence of questioning. Simply that questioning, in fact, plays a central role in facilitating the smooth exchange of epistemic goods within our epistemic communities. An epistemic community in which questioning did not take place would most probably be significantly more unstructured and fragmented than one in which it does. It is, I think, for precisely this reason that questioning features so prominently in our daily lives, successfully transcending otherwise deep-seated cultural and linguistic boundaries. Questioning is an indispensable form of social and epistemic cohesion.

This cohesion extends beyond the direct interactions of questioners and answerers in interpersonal exchange. Questioning also benefits the passive recipients of epistemic goods that have been acquired through the questioning of others. My knowledge of Wittgenstein's early life, for example, can be largely credited to the insightful questioning of his biographer. Much of the information gleaned from a news report is acquired through the skilful and selective questioning of journalists. One's decision to walk through the jungle might rest on the earlier results of Fred's local tiger survey. Indeed, many of the decisions we make throughout our lives, whether they are simple everyday choices or life-changing resolutions, are likely to be informed in some part, by epistemic goods acquired by others through questioning.

In this respect, then, the practice of questioning is partly constitutive of the epistemic community itself. In other words, the practice of questioning is part of what makes an epistemic community what it is: a group of individuals that produces, shares and consumes epistemic goods. The value of epistemic communities, and the role of questioning within them, can be emphasised once again by returning to Craig's (1999) genealogical account of knowledge. According to this account, knowledge ascriptions are a means of identifying good informants within an epistemic community. Fred has knowledge about the local tigers that I can benefit from. This is precisely the type of thing that makes being part of an epistemic community valuable. We can draw on the epistemic resources of others in order to improve our lives (and chances of survival) in myriad ways. As I have argued, questioning plays a central role in this story. It is because I am able to ask Fred whether he thinks there is likely to be a tiger in the area that I can benefit from the knowledge that he has and I don't. Again, questioning is an indispensable form of social and epistemic cohesion, one which helps substantially to form, sustain and grow our epistemic communities. This is the social virtue (or value) of questioning.⁴

6 The epistemology of questioning

I have argued that questioning is an epistemic practice and, moreover, one of significant social and societal value. If this is right, then questioning should be a topic of interest within contemporary epistemology, and perhaps in particular for social and virtue epistemologists. Invitations to write on the topic for collections such as this indicate that, at least to some extent, it is. Nonetheless, it is worth taking the final few paragraphs to emphasise the scope and rich potential of a contemporary epistemology of questioning, with the hopes of attracting new attention from epistemologists in the near vicinity and further broadening the field.

In the first instance, one can easily compare and contrast the epistemic practice of questioning with the closely related practice of testimony. As a topic of philosophical interest, testimony has received a great deal of attention in recent epistemology and has, in particular, occupied a significant portion of the social epistemology literature (Coady 1992; Matilal and Chakrabarti 1994; Fricker 2004; Lackey 2006; Lackey and Sosa 2006; Adler 2012). The pervasive nature of testimonial practices and their apparently central role in our epistemic communities justifiably make this a topic of pivotal concern for social epistemologists.

Much of the discussion has focused on the role of testimony in the sharing or transmission of epistemic goods, as manifested in the reductionism/anti-reductionism debate. It has even been suggested that the

very project of social epistemology rests on the outcome of this debate (Schmitt 1994; Goldberg 2010). In *Relying on Others*, for example, Goldberg (2010) emphasises the significance of testimony:

[I]t should come as no surprise that a project aimed at examining the anti-individualistic implications of our epistemic reliance on others should begin by taking an extended look at testimonial belief.
(Goldberg 2010, 11)

Goldberg goes on to quote Schmitt (1994) as claiming that testimony is “the most fundamental test of epistemological individualism” (Schmitt 1994, 4 quoted in Goldberg 2010, 11). As such, testimony has emerged as a practice of principal concern in contemporary epistemology.

Notably, the topic of testimony is often approached within social epistemology with a focus on the role or function of testimonial exchange in our epistemic communities. Broadly speaking, it is a mechanism for transmitting information.⁵ The role or function of questioning, I think, merits the same degree of attention. According to the genealogical account sketched above, questioning, like testimony, occupies a central role in our social epistemic interactions. It is a powerful and pervasive mechanism for eliciting information. If this is right, then it looks like questioning and testimony constitute two sides of the same coin, or are, at the very least, intimately related, performing complementary roles in our epistemic communities. Testimony functions to transmit information and questioning functions to elicit it.

Indeed, perhaps somewhat provocatively, one might view questioning as in some sense the more basic or fundamental of these practices, given that it is by asking questions that the transmission of information is often initiated. It is, for example, by asking Fred if he thinks there are any tigers on the prowl that I prompt him to testify, one way or the other. One could easily get into a chicken and egg debate about the relative primacy of questioning and testimony. I am not advocating for a particular position in that debate but rather making a case for the interest and significance of the debate itself, and the relationship between questioning and testimony more generally, within contemporary social epistemology. At the moment, it seems, we only have the egg and the picture would be at least more comprehensive if we also considered the chicken.

One might be tempted to push back and argue that, while questioning often leads to or initiates testimony, it is nonetheless testimony itself that demands attention within social epistemology, not the process that brings it about. Or perhaps we could go one step further and argue that it is testimonial belief, as Goldberg (2010) suggests, that requires our attention. In short, it is the epistemic product, not the process, that is of interest. By these lights, one might well appreciate the significant role

that questioning plays as an epistemic practice – one that regularly helps us to acquire epistemic goods – and yet fail to appreciate its import more generally for contemporary epistemology.

For this reason, it is worth making a brief case for the value of the process, as well as the product. Zagzebski's (2003) well-known example of a reliable coffee machine will be helpful here. This example was originally employed as a challenge to the reliabilist account of the value of knowledge. Zagzebski argued that the reliability of a coffee machine confers no extra value on the coffee it produces and, by analogy, that the value of a reliable belief-forming mechanism confers no extra value on the beliefs it produces. Whether or not that's right, it does not, of course, follow that the coffee machine itself deserves no further attention. In fact, quite the opposite. The fact that the coffee machine produces coffee – something that we value – suggests that we should pay close attention to its proper functioning and maintenance. This is precisely to ensure that it continues to produce the thing that we value, namely, coffee. Moreover, if it is not simply coffee that we are after, but *good* coffee, then we ought to pay extra attention to the machine that produces it in order to ensure that it produces just this. In the absence of a reliable coffee machine, we would regularly have no good coffee.

Analogously, the significant role that questioning plays in our epistemic lives, as a powerful and pervasive mechanism for eliciting information, gives us reason to pay close attention to the practice. Questioning often leads us to epistemic goods such as information, knowledge and understanding, and it facilitates the efficient exchange of these goods between us, often in the form of testimony. Without it, our epistemic lives would almost certainly be less rich, less informed and less interesting, and our lives, more generally, full of unexpected tigers in the jungle. Moreover, if it is not simply any old epistemic goods that we are after but particularly useful or illuminating ones, then we ought to pay extra special attention to the mechanism that leads us to them. In the absence of good, effective, timely questioning, we will regularly miss out on the most valuable epistemic goods, those that can profoundly enrich our epistemic lives and communities. As such, questioning provides an important subject for normative analysis in its own right.

The focus on good questioning, moreover, brings the significance of the practice for contemporary virtue epistemology to the fore. Or, perhaps more accurately (and in the spirit of the present collection), for social virtue epistemology. Questioning is an epistemic practice meaning, as I have argued, that it is a socially established set of activities directed toward the common epistemic goal of eliciting information. As such, questioning involves the coordinated efforts of groups of individuals, to form, sustain and grow their epistemic communities. In other words, if we want good questioning, we need good questioners. Skilled, well-motivated or even virtuous questioners, can and do contribute significantly to the epistemic

communities in which they live and work. I have argued elsewhere that good questioning also serves as a catalyst for the development of many, if not all, of the intellectual virtues, and for intellectually virtuous inquiry (Watson 2018). How we cultivate good questioning, then, is a topic of interest for any social virtue epistemologists concerned with the healthy and prosperous functioning of our epistemic world.

These are, I believe, just some of the topics that should spark interest, particularly for social virtue epistemologists, in a contemporary epistemology of questioning. Doing so, however, is just one of the (perhaps more incidental) aims of this chapter. More directly, I have sought to offer an account of questioning as an epistemic practice and, in doing so, an account of the social virtue of questioning. The genealogical route I have taken provides, I think, an illuminating insight into the social role and function of questioning, extending the study of questions and questioning beyond formal logical or linguistic analyses. This is just one route into a topic that is, I believe, of notable contemporary philosophical significance.

Notes

- 1 Interestingly, Roberts and Wood (2007) expand the list of potential goods at which an intellectual practice may aim to include the ‘powers and skills’ by which a person acquires epistemic goods. As such, one can engage in an intellectual practice, not only in order to acquire epistemic goods, but in order to improve one’s ability to acquire such goods.
- 2 Craig’s genealogical account has not been without criticism. In particular, critics have focused on the concept of ‘objectivisation’, which Craig introduces some way into the account (see, for example, Shapin 1994; Kelp 2011). The concept of objectivisation is not relevant for the present discussion so I will not explore these criticisms here.
- 3 To reiterate a point made earlier, this is not to say that questioning cannot take place outside of an explicitly social setting. Indeed, we often engage in questioning privately. Even when ostensibly performed outside of a social setting, however, questioning is still governed by the norms under which it operates within society. Just as meditation and prayer are frequently performed alone by individuals and yet still adhere to the norms of their practice within a wider community, so too does private questioning adhere to the norms of questioning in a social setting. If one diverges from these norms, then one cannot be said to be engaging in the practice at all. Thus, even questioning done in private is ultimately a socially established practice.
- 4 This is to say nothing of the significance of questioning as a social or interpersonal skill. We have probably all experienced people whose questions are too probing or insensitive or, on the other hand, people who don’t ask questions at all, even when it would be socially polite to do so (think of the common complaint that a partner on a date failed to ask any questions). The social virtue of questioning takes on a subtly different sense in the light of these types of examples and is an enticing topic for another time.
- 5 I do mean broadly here as there is much interesting debate about this in the contemporary literature, which I cannot hope to do justice to in the present context.

References

- Adler, Jonathan. 2012. Epistemological Problems of Testimony. In: *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta.
- Aloni, Maria. 2005. A Formal Treatment of the Pragmatics of Questions and Attitudes. *Linguistics and Philosophy* 28(5): 505–539.
- Alston, William. 1989. A ‘Doxastic Practice’ Approach to Epistemology. In: *Knowledge and Skepticism*, edited by Marjorie Clay and Keith Lehrer. Colorado: Westview Press, 1–29.
- Åqvist, Lennart. 1965. *A New Approach to the Logical Theory of Interrogatives*. Uppsala: University of Uppsala.
- Aristotle. 1984. *Metaphysics*. Translated by XXX. In: *The Complete Works of Aristotle: The Revised Oxford Translation*, Volume Two, edited by Jonathan Barnes. Princeton: Princeton University Press.
- Baehr, Jason. 2011. *The Inquiring Mind*. Oxford: Oxford University Press.
- Barnes, Barry. 2001. Practice as Collective Action. In: *The Practice Turn in Contemporary Theory*, edited by Theodore Schatzki, Karin Knorr Cetina, and Eike von Savigny. London: Routledge, 25–36.
- Belnap, Nuel, and Thomas Steel. 1977. *The Logic of Questions and Answers*. Connecticut: Yale University Press.
- Ciardelli, Ivano. 2010. A First-Order Inquisitive Semantics. In: *Logic, Language, and Meaning: Selected Papers from the Seventeenth Amsterdam Colloquium*, edited by Maria Aloni, Harald Bastiaanse, Tikitou de Jager, and Katrin Schulz. Berlin: Springer, 234–243.
- Coady, Cecil. 1992. *Testimony: A Philosophical Study*. Oxford: Oxford University Press.
- Craig, Edward. 1999. *Knowledge and the State of Nature*. Oxford: Oxford University Press.
- Cross, Charles, and Floris Roelofsen. 2018. Questions. In: *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. <<https://plato.stanford.edu/archives/spr2018/entries/questions/>>.
- Fricker, Elizabeth. 2004. Testimony: Knowing through Being Told. In: *Handbook of Epistemology*, edited by Ilkka Niiniluoto, Matti Sintonen, and Jan Wolenski. Dordrecht, The Netherlands: Kluwer, 109–130.
- Ginzburg, Jonathan. 1996. The Semantics of Interrogatives. In: *The Handbook of Contemporary Semantic Theory*, edited by Shalom Lappin. Basil: Blackwell, 385–422.
- Goldberg, Sanford. 2010. *Relying on Others: An Essay in Epistemology*. Oxford: Oxford University Press.
- Goldman, Alvin. 1999. *Knowledge in a Social World*. Oxford: Oxford University Press.
- Groenendijk, Jereon. 1999. The Logic of Interrogation. In: *Semantics and Linguistic Theory*, edited by T. Matthews and D. Strolovitch. Cornell: Cornell University Press, 109–126.
- Higginbotham, James. 1996. The Semantics of Questions. In: *The Handbook of Contemporary Semantic Theory*, edited by Shalom Lappin. Basil: Blackwell, 361–383.
- Hintikka, Jaakko. 1981. On the Logic of an Interrogative Model of Scientific Inquiry. *Synthese* 47: 69–83.

- Hintikka, Jaakko. 1983. New Foundations for a Theory of Questions and Answers. In *Questions and Answers*, edited by Ferenc Kiefer. *Linguistic Calculation* special issue, Cambridge: Cambridge University Press, 159–190.
- Jaworski, William. 2009. The Logic of How Questions. *Synthese* 166: 133–155.
- Karttunen, Lauri. 1977. The Syntax and Semantics of Questions. *Linguistics and Philosophy* 1(1): 3–44.
- Kelp, Christoph. 2011. What's the Point of "Knowledge" Anyway? *Episteme* 8(1): 53–66.
- Koura, Antti. 1988. An Approach to Why-Questions. *Synthese* 74: 191–206.
- Lackey, Jennifer. 2006. Knowing from Testimony. *Philosophy Compass* 1: 1–17.
- Lackey, Jennifer, and Ernest Sosa. (eds.) 2006. *The Epistemology of Testimony*. Oxford: Oxford University Press.
- MacIntyre, Alasdair. 1981. *After Virtue: A Study in Moral Theory*. Indiana: University of Notre Dame Press.
- Matilal, Bimal, K., and Arindam Chakrabarti, eds. 1994. *Knowing From Words*. Dordrecht: Kluwer Academic Publishers.
- Plato. 1997. *Theaetetus*. Translated by M. J. Levett and Myles Burnyeat. In: *Plato, Complete Works*, edited by John, M. Cooper. Indiana: Hackett Publishing Company.
- Prior, Mary, and Arthur Prior. 1955. Erotetic Logic. *Philosophical Review* 64: 43–59.
- Roberts, Robert, and W. Jay Wood. 2007. *Intellectual Virtues: An Essay in Regulative Epistemology* Oxford: Clarendon Press.
- Rouse, Joseph. 2006. Practice Theory. In: *Handbook of the Philosophy of Science Volume 15: Philosophy of Anthropology and Sociology*, edited by Dov M. Gabbay, Paul Thagard, John Woods, Stephen Turner, and Mark Risjord. Dordrecht: Elsevier, 630–681.
- Schatzki, Theodore (2001) Introduction: Practice Theory. In Schatzki, Theodore, Karin Knorr Cetina, and Eike von Savigny, eds. 2001. *The Practice Turn in Contemporary Theory*. London: Routledge.
- Schmitt, Frederick. 1994. Socializing Epistemology: An Introduction through Two Sample Issues. In: *Socializing Epistemology*, edited by Frederick Schmitt. Lanham, MD: Rowman and Littlefield, pp. 10–23
- Shapin, Steven. 1994. *A Social History of Truth: Civility and Science in Seventeenth-Century England*, Chicago: University of Chicago Press.
- Turner, Stephen. 2001. Throwing Out the Tacit Rule Book: Learning and Practices. In: *The Practice Turn in Contemporary Theory*, edited by Theodore Schatzki, Karin Knorr Cetina, and Eike von Savigny. London: Routledge, 129–139.
- Van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Oxford University Press.
- Watson, Lani. 2021. What Is a Question. *Philosophy: Royal Institute of Philosophy Supplements* 89: 273–297.
- Watson, Lani. 2018. Educating for Good Questioning: A Tool for Intellectual Virtues Education. *Acta Analytica* 33(3): 353–370.
- Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Oxford: Basil Blackwell Ltd.
- Zagzebski, Linda. 2003. The Search for the Source of Epistemic Good. *Metaphilosophy* 34(1/2): 12–28.

14b Commentary from J. Adam Carter

Lani Watson's 'The Social Virtue of Questioning: A Genealogical Account' offers a thoughtful and, on the whole, very plausible picture of questioning's place in social epistemology, where it is often overlooked. In Watson's view, questioning is best theorised about as an *epistemic practice*; it is socially established, activity based, and aimed at the epistemic end of eliciting information. This picture of questioning as an (epistemically aimed) practice is supported in part by a kind of Craigan (Craig 1991) genealogical strategy, one that is used, additionally, to support what is arguably the key thesis in Watson's chapter, which is that questioning is an indispensable form of social and epistemic cohesion, one which helps substantially to form, sustain and grow our epistemic communities. It is in this respect that the practice of questioning is meant to be understood as a *social* epistemic virtue.

While I am on board with Watson's wider picture here – I think there is a lot right about it – I am going to use this brief space to raise a few small quibbles, which I hope might prompt useful further discussion. These concern (i) the status of questioning as an epistemic practice; (ii) the methodology of the Craigan genealogical strategy, as applied to questioning; and (iii) the normative thesis Watson embraces about the social-epistemic value of questioning in an epistemic community.

First, regarding the status of questioning as an epistemic practice. It is a practice according to Watson because it meets three criteria; it is socially established, activity-based, and directed towards common goals; it is an *epistemic* practice because the common goal is an epistemic one, that of eliciting information. As Watson rightly points out, *sometimes* questioning is used to serve non-epistemic purposes. We might question someone in order to undermine their authority; or to distract, to impress, to show off, to put someone on the spot, to 'shoot the bull', uninterested in eliciting information. The fact that questioning is often used *for* these ends, however, should lead us to ask why the aim of eliciting information is privileged among these ends such that *questioning* is best understood as an epistemic practice – as opposed to – a wider multifariously aimed practice that on occasions is epistemically oriented.

Perhaps an answer here might come from Watson's Craigian genealogical story: if a society existed without questioning, we'd need to invent it in order to serve the valuable function of eliciting information. But this kind of answer leads to my second critical point, which concerns Watson's use of the Craigian strategy. Showing off, impressing, undermining authority, shooting the bull – these are also valuable – and questioning is a mechanism by which we can do all of this quite directly. Which raises a kind of 'devil's advocate' question for Watson's Craigian strategy: might not a society without a mechanism that could do all of *these* things so effectively be led to embrace questioning as a flexible way to facilitate all of these goals? Of course, one might point out that you can achieve these other goals by mechanisms other than questioning. True, but by the same token, you can elicit information by means other than questioning – viz., including via command or threat. This is not to say that Watson's Craigian strategy is implausible; rather, that a more refined version of it might help us to better connect (as Watson wants to) the practice of questioning with the epistemic function of eliciting information.

A third place where I'd like to press Watson's argument concerns the *social value* of questioning as a practice. Here I will quote a key passage I'd like to critically discuss:

[...] the value of questioning from a social, or perhaps societal, perspective is not hard to see. Questioning allows us to both access and generate epistemic goods such as true belief, justification, information, knowledge and understanding – goods that we value within our epistemic communities. Equally, if not more importantly, questioning facilitates the exchange of these goods among community members – the *sharing* of information, knowledge and so on. As such, questioning serves to ease the passage of epistemic goods between members of an epistemic community. Much as the holding open of doors eases passage in and out of buildings. Indeed, it is hard to see how the smooth and efficient exchange of epistemic goods would be possible in the absence of questioning.

In the above passage, Watson is defending the social value of questioning by drawing our attention to various ways in which questioning is instrumentally epistemically valuable in an epistemic community. As she rightly points out, questioning allows us to access and generate epistemic goods, it facilitates the sharing of information, knowledge, and so on.

Given that the reasoning here is instrumental reasoning, it is fair enough to ask: what is the *nett* instrumental epistemic value of questioning in a community? To Watson's credit, it is probably positive. However, it should at least be registered that the practice of questioning can generate all of these goods only by at the same time putting us at

epistemic *risk*; the practice of questioning puts us at risks of misinformation, deception, and betrayal; these risks are inevitably incurred by questioning aimed at eliciting information.

An epistemic doomsayer might then suggest the following kind of counter-reasoning, in response to Watson's optimistic passage above:

Questioning allows us to both access and generate epistemic *bads* – viz., misinformation, false beliefs, etc. – that we disvalue within our epistemic communities. Equally, questioning facilitates the exchange of misinformation among community members – by facilitating the *sharing* of misinformation. As such, questioning serves to ease the passage of misinformation between members of a community. Much as holding doors eases passage in and out of a building. Indeed it is hard to see how the smooth and efficient exchange of misinformation would be possible in the absence of questioning.

The above is the reasoning of the doomsayer. Watson's is the reasoning of the optimist. My own thinking here is less committal and more curious. I think Watson has a point, but so does the epistemic doomsayer. My final question to Watson is: why throw in with the optimist? And relatedly, *can* we defend the optimist here without inadvertently reducing the value of questioning in a social community to the value of trust?

Reference

Craig, Edward. 1991. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Clarendon Press.

14c Commentary from S. Goldberg

Interrogative Attitudes and the Social Practice of Asking Questions

It is a curious feature of contemporary philosophy that, until very recently, there had been little attention devoted to questions, and almost no attention to them outside of certain parts of formal semantics. Lani Watson deserves significant credit for the role she has played in getting the profession to rectify this situation. Watson's present chapter is yet another example of these efforts. I am deeply sympathetic with Watson's overarching aim to get philosophers (and epistemologists in particular) to attend to questions, and I embrace her idea that questioning itself is fruitfully conceived as a social practice. In this brief commentary, I will focus on one claim that is advanced as part of her 'genealogical account' of that practice. The claim in question is an explanatory hypothesis regarding the relationship between questioning as a social practice, and the role questions play (or can play) in more 'private' settings.

Stated programmatically, Watson's own position appears to be this. A genealogical account of the practice of questioning, aimed at discerning why such a practice might have arisen in a community in which it didn't already exist, can illuminate the point of the practice (the purpose it serves). What such an account shows, she argues, is that the practice is constituted by "*a socially established set of activities directed towards the common goal of eliciting information*" (MS, 9; italics in original). This makes the practice an epistemic one, and hence one worthy of the attention of social epistemologists. She acknowledges that there are activities that involve questions that don't aim at this goal: there are rhetorical questions, for example, and also 'private' questions that one puts to oneself. But these can be regarded as derivative phenomena whose features can be explained by reference to those of the core phenomenon itself: 'activities directed towards the common goal of eliciting information'.

It is regarding this last point that I want to express some doubts. The target of my doubts is a specific view Watson expresses in an extended

footnote. Having presented her genealogical account focusing on the *social practice* of questioning, Watson writes the following (footnote 3, p. 9 of the MS):

... this is not to say that questioning cannot take place outside of an explicitly social setting. Indeed, we often engage in questioning privately. Even when ostensibly performed outside of a social setting, however, questioning *is still governed by the norms under which it operates within society*. Just as meditation and prayer are frequently performed alone by individuals and yet still adhere to the norms of their practice within a wider community, so too does private questioning adhere to the norms of questioning in a social setting. If one diverges from these norms, then one cannot be said to be engaging in the practice at all. Thus, *even questioning done in private is ultimately a socially established practice*.

(MS, p. 9; italics added)

Watson's position here seems to be that the social practice of asking questions is *explanatorily fundamental* in the following sense: when it comes to questions that take place in "private" settings, the relevant features of the phenomenon can be explained in terms of the norms that govern the social practice of questioning.

As I say, I have my doubts. I think that there is a case to be made that the situation is exactly the reverse: the social norms of questioning themselves emerge out of and reflect the fundamental role that questions play in a more 'private' setting – in inquiry. Here I have in mind the role in inquiry of what Jane Friedman has called the 'interrogative attitudes'.¹ These are the *question-directed attitudes* – attitudes like wondering, being curious – which, according to Friedman, constitute the sort of attitude proper to the activity of inquiring (and which have a question as their content). For example, consider the detective who wonders who committed the murder, or the teacher who is curious whether Sam passed the exam. On the assumption that such interrogative attitudes do in fact have questions as their contents (as per Whitcomb 2010; Friedman 2013; Carruthers 2018), Watson's approach implies that the norms governing these "private" questions just are the norms that govern the social practice of asking questions.²

But there are grounds for doubt. These come in the form of (at least) two reasons to think that the explanatory relation runs in the reverse direction. First, as Carruthers (2018) emphasises, non-human animals exhibit attitudes like curiosity (as well as the behaviours that such attitudes motivate). In light of this, we might reasonably expect that a theory of questions and questioning should explain how the social activity of questioning might have emerged out of more basic forms of questioning (or proto-questioning). Second, it seems plausible to think

that the standards on the social activity of questioning aim (at least in part) at ensuring sensitivity to the epistemic perspective of one's target audience, whereas such sensitivity is simply irrelevant to cases in which one is posing a question to oneself. Wondering about something can be perfectly proper (in any relevant sense of 'proper') even under conditions in which questioning another person (in an attempt to elicit an answer) is not – say because it is common knowledge that the answer is not known. This suggests that, far from seeing the 'private' case as following norms already present in the social practice of asking questions, we do better to think that the latter emerged out of the former, under the social pressures that render us sensitive to the epistemic perspectives of others.

It seems, then, that there are some reasons for thinking that questions play an important role in individual inquiry and that the social practice of questioning might be usefully understood as emerging out of that role. None of this is meant to diminish the importance of the social practice of questioning, let alone to deny that such a practice is governed by norms of a sort in which social epistemologists ought to take an interest. On the contrary, it is a credit to Lani Watson's important work in this area that it puts us in a position to raise these issues connecting the social practice of questioning with the question-directed attitudes to which Friedman and others have drawn our attention. I celebrate that even if I harbour doubts about the particular explanatory hypothesis she advances here.

Notes

- 1 See e.g. Friedman (2013, 2017, 2019). This way of thinking about questions goes back at least to Rescher (2000), Hookway (2008) and Whitcomb (2010), and it has been usefully explored too in Carruthers (2018).
- 2 To repeat, Watson says, "Just as meditation and prayer are frequently performed alone by individuals and yet still adhere to the norms of their practice within a wider community, *so too does private questioning adhere to the norms of questioning in a social setting*" (italics added).

References

- Carruthers, Peter. "Basic questions." *Mind & Language* 33.2 (2018): 130–147.
- Friedman, Jane. "Question-directed attitudes." *Philosophical Perspectives* 27.1 (2013): 145–174.
- Friedman, Jane. "Why suspend judging?" *Noûs* 51.2 (2017): 302–326.
- Friedman, Jane. "Inquiry and belief." *Noûs* 53.2 (2019): 296–315.
- Hookway, Christopher. "Questions, epistemology, and inquiries." *Grazer Philosophische Studien* 77.1 (2008): 1–21.
- Rescher, Nicholas. *Inquiry dynamics*. Transaction Publishers, 2000.
- Whitcomb, Dennis. "Curiosity was framed." *Philosophy and Phenomenological Research* 81.3 (2010): 664–687.

14d Lani Watson's Response to Commentaries

Goldberg and Carter raise a set of distinct yet insightful challenges in response to my chapter. I thank them both and address their comments in turn. First, Goldberg poses an intriguing challenge regarding the explanatory relation between the social practice of questioning and its private counterpart. He notes that I elaborate on this in a footnote, where I indicate that even when questioning takes place outside of a social setting, the norms that govern it are those derived from the social practice, rather than the other way around. Goldberg puts pressure on this claim, suggesting two reasons for taking the explanatory relation to be the exact reverse.

Goldberg says that “a theory of questions and questioning should explain how the social activity of questioning might have emerged out of more basic forms of questioning (or proto-questioning)”. This seems right and indeed the notion of ‘proto-questioning’ is one that I have constructed and deconstructed for my own purposes at times, with a view to attempting precisely this kind of explanation. It would no doubt enhance any theory of questioning to understand precisely when and how it features in the early stages of cognitive development and it seems plausible to me that one of these stages could be termed ‘proto-questioning’.

Importantly, however, this proto-questioning is not, from my perspective, questioning proper. This matters in relation to Goldberg’s challenge because one might construe proto-questioning as a paradigmatically private activity; think of a baby reaching out to discover the texture of a shiny material. It is unclear whether such an act should be construed as a question, but it is surely part of the process by which babies learn to question. This is what I have in mind when I think of proto-questioning. Significantly, the transition from proto-questioning to questioning plausibly relies, in part, upon recognising that *others can provide information* that one does not have. This social dimension is what transitions proto-questioning into questioning proper and, as such, the norms that govern questioning are social, rather than private norms.

This social dimension is also significant in the case of wondering. Goldberg notes that “Wondering about something can be perfectly proper...even under conditions in which questioning another person (in

an attempt to elicit an answer) is not". The switch between wondering and questioning here is important because, much like proto-questioning, wondering is not, from my perspective, questioning proper. It is a state that often leads to questioning; in Friedman's (2013) terms, it is an interrogative attitude. But this state or attitude is not itself questioning. Again, this matters because, like proto-questioning, one might construe wondering as a paradigmatically private activity. I often wonder to myself about the scale and nature of the universe, but it is unclear whether such an act should be construed as a question.¹ As such, while I agree with Goldberg that *wondering* can be perfectly proper 'even under conditions in which questioning another person is not', it is less clear that private *questioning* can be. If I know that the answer to my question is not known, by me or anyone else, it certainly seems odd to ask a question in an attempt to elicit an answer, whether from myself or anyone else. By contrast, I can wonder to my heart's content.

Moreover, Goldberg draws on the idea that "the standards on the social activity of questioning aim...at ensuring sensitivity to the epistemic perspective of one's target audience". I am not convinced that the sensitivity Goldberg is talking about is 'simply irrelevant', as he contends, in the case of private questioning. Despite a longstanding curiosity concerning the nature of the universe, I am not an astronomer or physicist and so have little in-depth knowledge when it comes to answering questions about its scale. Again, I can, and often have *wondered* about the size of the universe, but it makes little sense to ask myself the question 'how big is the universe', in an attempt to elicit that information. This is precisely because I am duly sensitive to my own limited epistemic perspective on this score.

The general point being that the constraints or norms that govern the public practice of questioning appear to apply equally in the private case: it is not obvious that I can properly pose a question to myself that I could not properly pose to another person, and vice versa. These social norms do not apply in cases of proto-questioning or wondering, whether public or private, because these are not cases of questioning. Rather, they represent the borders of the practice, and so serve to define it. All that said, I take Goldberg's challenge as a welcome invitation to explore the concepts of proto-questioning and wondering in greater depth. This is something that I hope to do in future work, with a view to explaining the relationship between these concepts and questioning proper. Understanding this relationship will be an important piece of the puzzle when developing an in-depth genealogical account of questioning.

In his response, Carter raises the spectre of the 'epistemic doomsayer'; one who perceives the function of questioning in terms of the easy exchange of *misinformation* (as opposed to true information) in an epistemic community. Carter characterises my position, by contrast, as one of optimism with respect to the function of questioning and presses for

a more substantive defence of this optimistic stance with a pair of excellent and provocative questions.

I contend that it is misleading to characterise my position as one of mere optimism. As Carter rightly identifies, the genealogical account of questioning is intended to serve as the (or at least a) basis for the claim that the primary function of questioning is information elicitation. The genealogical story is meant to provide a plausible answer to the genealogical question of why a society would develop the practice of questioning in the first place. Much like one can ask why prehistoric humans developed hammers and hammering. A hammer can be used as a paperweight or to prop open a door but these uses do not, in themselves, provide a plausible answer to the question of why humans developed hammers and hammering. The genealogical story is meant to provide a *plausible* answer to the genealogical question. The account I offer is, I think, more plausible than that of the epistemic doomsayer because questioning gets us something *that we value*. It is the fact that we value the thing that we get that makes the optimistic account more plausible than the doomsayer account. These accounts are not on an equal footing if one takes the genealogical framing seriously.

Nonetheless, Carter is right to highlight the epistemic risks associated with questioning: “risks of misinformation, deception, and betrayal”. This is a topic worthy of much greater attention. Furthermore, Carter pushes for a more refined version of the genealogical story (also implicit in Goldberg’s commentary) and I am grateful for this push. The Craighian inspiration for the genealogical story comprises an extensive body of work and a project on a similar scale is required in order to develop and refine the genealogical account of questioning. My chapter serves as a starting point for such a project and, crucially, aims to provide some motivation for embarking upon it. Specifically, I aim to highlight the significance of questioning as an epistemic practice for social and virtue epistemologists. As I read both Goldberg’s and Carter’s constructive comments, the chapter has been broadly successful on this score.

Note

- 1 I discuss this precise example in Watson (2021), where I examine the results of a large survey and conclude from the results that wonderings such as this represent a significant and fascinating grey area in the analysis of questions.

References

- Friedman, Jane. 2013. Question-directed attitudes. *Philosophical Perspectives* 27(1): 145–174.
- Watson, Lani. 2021. What is a question. *Philosophy, Royal Institute of Philosophy Supplements* 89: 273–297.

Part IV

**Methods and
Measurements**

T&F Proofs – Not for Distribution

T&F Proofs – Not for Distribution

15 An Interdisciplinary Methodology for Studying Collective Intellectual Character Traits

T. Ryan Byerly

This chapter will describe an interdisciplinary methodology for studying collective intellectual character traits. The methodology described—especially its empirical component—is not the only possible or potentially valuable methodology for studying collective intellectual character. Nor will I argue that it is superior to other methodologies, though it should be clear from the chapter that this approach does have the benefit of being rather simple. The main aim of the chapter is the more modest one of articulating what is involved in employing this methodology in the hopes that it might be experimented with more widely to see whether it can advance our understanding of collective intellectual character.

At a very abstract level, the methodology has two main components: a conceptual component and an empirical component. Researchers first conceptualize the collective intellectual character traits of interest. Then, they operationalize these traits and collect and analyze data about particular collectives in order to ascertain the relationships between the traits they have conceptualized and other variables of interest. I begin, accordingly, by describing the conceptualization of particular collective intellectual character traits in Section 1. In Section 2, I explain how collective intellectual character traits can be operationalized and studied using a method akin to that which has been used to study organizational climate and organizational virtuousness.

1 Conceptualizing Collective Intellectual Character Traits

The conceptual component of the proposed interdisciplinary methodology focuses on conceptualizing particular collective intellectual character traits. To conceptualize particular collective intellectual character traits well, researchers need to have an idea of what collective intellectual character traits are more generally. Having an idea of what these are serves both to illuminate the range of traits that might be conceptualized as well as the features of traits of interest that may need explication. So, I will start in this section by addressing the question of what collective intellectual character traits, in general, are.

The simple answer is that collective intellectual character traits are collective, intellectual, characterological, and trait-like. Each of these components can be explained more fully. I'll go in reverse order.

What makes the relevant features traits is that they are unified tendencies to display a wide range of characteristic behaviors under a wide range of characteristic triggering circumstances. The individual personality trait of openness to experience, for example, is a tendency to display openness toward new experiences (McCrae and Costa 1997). People who are highly open to experience tend to notice when opportunities for gaining new experiences arise. And, when they detect an opportunity to have a new experience, they tend to greet the opportunity with positive emotions and judgments and a willingness to try it out. In this way, full-blow traits involve tendencies of emotion, cognition, perception, and volition.

What makes the traits that are our focus characterological is that they reveal their possessor's values or motives (cf. Battaly 2015, 19). For a trait to be a character trait, the unified tendencies of the trait must be explained by unifying motives or values. The possessor of the trait tends to engage in these behaviors under the relevant circumstances because they possess the relevant values. If there are traits that do not reveal their possessor's values in this way, as some philosophers have suggested (cf. Battaly 2017, 678; Miller 2014, 9–18), then these traits wouldn't be character traits. They wouldn't reveal who their possessor is in the way that character traits distinctively do.

What makes the character traits that are our focus intellectual is that the motives or values they reveal are intellectual motives or values. There is a wide variety of such motives. Some epistemic agents are motivated to reach decisions quickly and stick to them (Webster and Kruglanski 1994). Others are motivated to secure good intellectual reputations for themselves (Roberts and Wood 2007, 236). Some are motivated to avoid revealing their ignorance (Tanesini 2018). Others are motivated to base their views on the best available evidence. Some are motivated to lead others to share their views (Saucier and Webster 2010). Others are motivated to promote others' epistemic well-being (Byerly 2021). All of these, and many others, are intellectual motivations. When an epistemic agent has a unified tendency to display a wide range of affective, cognitive, perceptive, and volitional behaviors out of motivations of these sorts, they have an intellectual character trait.

Finally, what makes the intellectual character traits collective is that they are possessed by groups (cf. Lahroodi 2019). They are traits that are sensibly attributed to groups, regardless of how the sensibleness of these attributions is best explained. Thus, for example, if a group of educators is highly sensitive to opportunities to enhance students' epistemic well-being, greets opportunities to improve students' epistemic well-being with positive emotions and judgments, and tends to make efforts

to advance students' epistemic well-being when opportunities arise, all because they value students' epistemic well-being, this would be an example of a collective intellectual character trait. To put it all together, collective intellectual character traits are tendencies of groups to display a wide range of affections, cognitions, perceptions, and volitions out of unifying intellectual motivations.

Notably, this account of collective intellectual character traits stands in parallel to an account of intellectual character traits of individual people. According to the latter approach, intellectual character traits of individuals are tendencies of these individuals to display a wide range of affections, cognitions, perceptions, and volitions out of a unifying intellectual motivation (cf. Baehr 2011).¹ An individual educator, for example, might be highly sensitive to opportunities to enhance students' epistemic well-being, greet opportunities to improve students' epistemic well-being with positive emotions and judgments, and tend to make efforts to advance students' epistemic well-being when opportunities arise, all because they value students' epistemic well-being.

This parallelism between the proposed account of collective intellectual character traits and the foregoing account of intellectual character traits of individuals raises the question of the exact relation between the two—a question that has been a focal point of interest for philosophers working on collective character traits (cf. Lahroodi 2019). When a collective intellectual character trait exists in a group, does it always exist only because the members of the group themselves possess this intellectual character trait? Summativists answer “yes”; antisummativists answer “no”.

The trend in philosophical work on collective character has been toward antisummativism. The main kind of argument given in defense of antisummativism appeals to cases in which group members tend to behave in a markedly different way in the group context than they would outside of it (see Lahroodi 2019 for a review). In these examples, a group appears to display a character trait while the group members in their private lives appear not to display it, or a group appears not to display a character trait, though its members do appear to display it in their private lives. Often, what plays a key role in these examples are the group's policies or procedures, whether formal or informal. The group has adopted policies or procedures that regulate their members' conduct when acting as group members, and these policies and procedures lead the group members to behave differently as group members than they would as private individuals. The group tends to display (or not) a unified range of behaviors because of group values encoded in the group's policies and procedures, but the individual group members, because they may not endorse these same values equally as private individuals, govern their private conduct differently. This might lead, for example, to racist groups composed of nonracist individuals, or the opposite.

A less well-known argument for antismmativism is also worth identifying here. This argument focuses on cases in which a group appears to manifest a character trait that just isn't available as a character trait for individuals, because of differences between groups and individuals (see Byerly and Byerly 2016). The most obvious example of a relevant difference between groups and individuals is that groups have members who may interact in the group's intellectual activities, whereas individual inquirers do not. As such, if there are any intellectual character traits concerned specifically with the regulation of group member interaction in group intellectual activity, these may be good candidates for distinctive group intellectual character traits that cannot be possessed by individual inquirers.

Two examples of such traits come to mind, but there is ample room for further exploration of this topic. First, one of the key, distinctive group intellectual activities is the distribution of intellectual labor (see Bird 2014). As such, we might think that there are group intellectual character traits focused upon the distribution of intellectual labor. What is involved in distributing intellectual labor virtuously in groups? I don't have a fully worked-out answer to offer, but presumably, any fully worked-out answer will want to include the group's commitment to distributing intellectual labor in a way that promotes its achievement of group intellectual aims, but that also balances this commitment with a commitment to the intellectual well-being of the group's members. The group that divides intellectual labor excellently will be skilled in identifying ways that intellectual labor can be divided, skilled in identifying the intellectual strengths and weaknesses of its group members, and skilled in matching its members to fitting portions of the divided group labor (cf. De Bruin 2015). The group will tend to exercise these skills in a way that is governed by motivations to achieve group epistemic goods, and that balances the achievement of these goods with the promotion of epistemic goods for group members.

A second example focuses on the group's activities in empowering (or disempowering) group members to contribute to group inquiry. To contribute well to group inquiry, group members may need to be provided with access to relevant materials, may need training in task-relevant skills, and may need channels of communication whereby they can appropriately influence group inquiry. A group that is excellent at empowering its members to contribute to group inquiry will be attentive to the needs of its group members and motivated to meet these needs so as to advance group inquiry.

Both of these arguments for antismmativism contain an important lesson for the project of operationalizing collective intellectual character. They both teach us that it will not always work to operationalize a group's possession of an intellectual character trait as a summation of group members' individual possession of this trait. We can't always just

assess whether the group members privately possess the trait of interest, and then reliably draw a conclusion on this basis about the extent to which the group possesses it. In some cases, this will not work because, while both the group and its members can possess the trait, there is a divergence between the group's possession of it and the members' possession of it. In other cases, this will not work because only the group and not its members can possess the trait.

So far, I have only offered an account of what collective intellectual character traits are. But it will also be instructive to consider what makes a collective intellectual character trait a virtue or a vice, or something in between. Roughly speaking, the virtues are the character traits that surpass a certain threshold of goodness, while the vices are the character traits that exceed a certain threshold of badness, and character traits between these thresholds are "mixed traits" (Miller 2014). The question here, though, is what the relevant sort of goodness or badness consists in when it comes to collective intellectual character traits. What sort of goodness is it that contributes toward making a collective intellectual character trait a virtue? What sort of badness is it that contributes toward making it a vice?

In the case of virtues and vices of individual people, there is a widely accepted answer to this question. What makes a character trait of an individual person a virtue is that it makes them better as a person; what makes it a vice is that it makes them worse as a person (cf. Battaly 2015, 5). We might debate exactly what it is to become better or worse as a person, but at least this much is commonly agreed.

Yet, it does not seem that this answer to the question transfers very well from the case of individual people to the case of groups. That is, it doesn't seem that the best approach to explaining what makes a collective trait a virtue is that it makes the group that possesses it better as a person. It may be that groups are sometimes appropriately treated as persons, at least for legal purposes. But, even still, it seems that the improving groups as persons is not what makes a collective trait a virtue. When we evaluate the characters of individual people, we do so with reference to the kind, person. We use our evaluations of their characters to judge how good or bad a person they are. But, when we evaluate the characters of groups, we don't do so primarily with reference to the kind, person. We don't use our evaluations of their characters primarily to judge how good or bad a person the group is. Instead, we use these evaluations to judge how good or bad they are in another respect.

This other respect isn't just their goodness or badness as a generic group, either. We don't primarily use our evaluations of the characters of groups to inform our judgments of how good they are as a generic group any more than we use them to judge how good they are as persons. The reason for this is that there are very different kinds of groups. While there may be some qualities of character that would make just any group

better as a group regardless of the kind of group that it is, many of the qualities of character that we care about in groups are not like this. Instead, they are qualities that make a group better as the particular kind of group that it is, and not merely better as a group in general or better as a person.

Thus, the kind of goodness that is relevant to collective virtues, including collective intellectual virtues, would seem to be of this sort (cf. Byerly and Byerly 2019). What makes a collective intellectual character trait a virtue is that it makes the group that possesses it better as the kind of group that it is. Likewise, what makes a collective intellectual character trait a vice is that it makes the group that possesses it worse as the kind of group that it is. This observation has important implications for studying collective intellectual virtues and vices. There may be different traits that are collective intellectual virtues or vices for different groups. A trait that is a collective intellectual virtue or vice for one group may not be a collective intellectual virtue or vice for other groups. Whether a trait is a virtue or vice for a group depends upon what sort of group it is, and whether possessing this trait makes it better or worse as that kind of group.

As a simple illustration, we might compare two different institutions of higher education with differing missions. One places a high priority on staff research with comparatively lower priority on teaching undergraduate students, and the other places a high priority on teaching undergraduate students and almost no priority on staff research. There are in fact many institutions of higher education that differ from one another in precisely these respects (Cummings and Shin 2014). The first we might call a research-focused institution, and the second a teaching-focused institution. Plausibly, they are different kinds of institutions, and different collective intellectual virtues will make them better or worse as the kinds of institutions they are. The kind of collective intellectual character trait focused on fostering students' intellectual well-being that we briefly described earlier will be very important for the second institution, but comparatively less important for the first institution. There may even be sub-groups within the first institution where this trait would not be a virtue at all, as these sub-groups may be devoted exclusively to research. The collective intellectual character traits that would be virtues for such a research-only subgroup would differ significantly from those that would be virtues for the teaching-focused institution.

Differences of this sort can be even more dramatic. After all, despite their differences, the imagined institutions in the previous example are still both institutions of higher education. As such, there may still be quite a bit of overlap between the traits that would be collective intellectual virtues for them. Yet, we could also contrast these institutions with other institutions that are even more different from them. For example, we might consider which traits would be intellectual virtues for

a troupe of comedians, or for a religious congregation. A witty tendency to detect, appreciate, and satirize each other's vulnerabilities may be a collective intellectual virtue for the comedy troupe but not the other groups; and a tendency to prioritize remembrance of a certain foundational religious message may be a collective intellectual virtue of the religious congregation but not the other groups. Very different traits make these groups better groups of their kinds because the groups are of very different kinds.

In conceptualizing collective intellectual character traits, virtues, and vices in the way proposed, I have attempted to be fairly ecumenical. I have only offered some basic parameters for thinking about what these traits, virtues, and vices are. There are many details I have left unspecified. Researchers may disagree, for example, about whether all traits are character traits. They may disagree about the precise elements that constitute character traits, virtues, or vices. They may disagree about what makes a motive an intellectual motive. They may disagree about how best to sort groups into different kinds. Even still, despite the possibility for disagreement about these details, the conception of collective intellectual character traits, virtues, and vices developed here should be agreeable to many researchers.

These conceptualizations should also be sufficient to provide important practical guidance for the project of conceptualizing particular collective intellectual character traits. In conceptualizing particular collective intellectual character traits, we should be guided by an understanding of the kinds of collectives we are interested in studying, and how different character traits would influence their quality as the kinds of groups they are. Guided by a conception of the nature of these groups, we can hypothesize about specific intellectual motivations that could unite tendencies of these groups to display a wide range of emotions, cognitions, perceptions, and volitions. And we can identify particular patterns of emotion, cognition, perception, and volition that would be characteristic of these unifying motivations. In this way, we would arrive at a fairly detailed and well-developed conceptualization of a particular collective intellectual character trait to study. If the hypothesized trait would make the groups in question sufficiently better as the kind of group they are, the trait is a candidate for being a collective intellectual virtue for these groups. If it would make the groups in question sufficiently worse as the kind of group they are, it is a candidate for a collective intellectual vice of these groups.

2 An Empirical Approach to Studying Collective Intellectual Character Traits

Once researchers have developed a conceptualization of a particular collective intellectual character trait, they may wish to study the role this trait

plays in particular groups. They may be interested, for example, in questions about how relevant groups that possess this trait in larger measure differ from relevant groups that possess it in lesser measure. They may be interested in antecedents of the trait—what might lead a group to be more strongly characterized by it or more weakly characterized by it. They may be interested in the consequences of the trait—which other features of the group and of other entities that interact with the group may be impacted by the presence or absence of the trait. For example, researchers might take an interest in what leads research teams to divide intellectual labor well, and how teams' tendencies to divide intellectual labor well influence the group's performance and features of the well-being of its members and of other individuals and groups with which the group interacts.

In this section, I will outline a methodology for conducting this kind of empirical study of collective intellectual character traits. The methodology is one that has already been employed fruitfully in the study of various "climates" of organizations. It has also been used to study various dimensions of virtuousness in organizations. The purpose of describing it here, then, is not to champion something entirely novel. It is instead to help advance wider understanding of the method and, combined with the work of the previous section, to illuminate how it might be used in the study of collective intellectual character in particular.

In many respects, the methodology mirrors the common methodology of using self-report questionnaires to study the character traits of individuals—as, for example, in the case of the widely used Values In Action Inventory of Character Strengths (Peterson and Seligman 2004). In using self-report questionnaires to measure the character traits of individuals, researchers ask individuals to respond to items about their own typical patterns of emotion, cognition, perception, and volition; their responses are assigned consistent mathematical values; and a score can be computed for each respondent for the trait in question. The score represents how "high" or "low" this individual is with respect to the focal character trait.

Items in the questionnaire reflect the researchers' conceptualization of the character trait in question. The items used are usually selected through a process that involves drafting a large original pool of items and narrowing this item pool through the use of statistical techniques such as exploratory and confirmatory factor analysis.² Often, self-report questionnaires produced through this type of method will include between four and fifteen items per trait. Ideally, the items included in a final questionnaire exhibit strong properties of reliability, such as a high Cronbach's alpha, high item-scale correlations, and high test-retest correlations. Also ideally, evidence for the questionnaire's validity can be obtained from its convergence with other measures to which researchers would expect it to be similar, or from its divergence from measures from which they would expect it to differ. Researchers can compare the scores

of individuals on the questionnaire with other variables of interest, such as participants' well-being or health. They can study the effectiveness of interventions designed to enhance the presence of the relevant character traits, and they can conduct longitudinal studies examining how changes in an individual's possession of a trait impact other variables of interest. Introductory texts such as (Furr 2011) describe the steps of these processes in detail.

It is worth observing here that some interdisciplinary research of this kind has been conducted which focuses specifically on intellectual character traits of individuals as conceptualized above. A good example of this is the research on intellectual humility first reported in (Haggard et al. 2018). In this work, a new scale was developed for measuring intellectual humility that was explicitly guided by the conceptualization of intellectual humility defended by a group of philosophers (Whitcomb et al. 2017). The philosophers had developed a detailed conception of the nature of virtuous intellectual humility in accordance with the above conception of intellectual virtues, which allowed for "specific predictions about the kinds of behaviors, motivations, and feelings that an intellectually humble person would demonstrate" (Haggard et al. 2018, 185). A team of philosophers and psychologists used this conceptualization to guide their work as they drafted a large pool of items to measure intellectual humility and then used exploratory and confirmatory factor analysis to create a shortened, three-factor scale to measure it, and sought evidence of the reliability and validity of the new scale. The final scale included items representing limitations owning (e.g., "When someone points out a mistake in my thinking, I am quick to admit that I was wrong"), love of learning (e.g., "When I don't understand something, I try hard to figure it out"), and appropriate discomfort with intellectual limitations (e.g., "I tend to get defensive about my intellectual limitations and weaknesses", reverse scored). In subsequent research, this kind of virtuous intellectual humility has been found to be associated with possessing more general knowledge, and with being more open-minded, curious, and reflective (Krumrei-Mancuso et al. 2020).

To follow a similar model in studying the character traits of groups, researchers would need for participants to complete questionnaires about the typical patterns of emotion, cognition, perception, and volition of groups of interest. The items used in such questionnaires would be about the patterns of behavior of the group and not of the individual completing the questionnaire. Researchers would need to assign mathematical values to the possible responses to the questionnaire in a consistent manner. A common approach used in research on individual character traits that could be replicated here would employ a Likert-scale anchored by "strongly disagree"/ "strongly agree" or "very much unlike us" / "very much like us". These mathematical values could be used in at least two different ways for research.

First, researchers can study the relationships between individual participants' perceptions of group character traits and other variables of interest. Here it is each individual participant's evaluations of the group that are compared to other variables. This approach does not require that participants in a study are members of the same group. Research of this kind can reveal ways in which individuals' perceptions of the collective intellectual character traits of groups of certain kinds are related to other variables of interest. For example, this kind of research could address how employees' perceptions of the intellectual character traits of their organizations are related to their own motivations or behaviors at work.

Second, researchers can aggregate the responses of multiple participants who are members of the same group using a direct consensus model (Chan 1998) and compare these aggregated values to variables of interest. In this case, the aggregated responses of multiple group members are used to determine a score for the group itself, and it is this group score that is then compared with other variables. Here it is the shared perception of group members that is in focus. This shared perception itself serves as a measure of the group's character, in roughly the same way that an individual's perception of their character serves as a measure of their character in the case of individual self-reports. This kind of research can address questions about how groups' intellectual character traits are related to other variables of interest. For example, it can address how changes to a group's policies affect the group's intellectual character, and it can address how groups' intellectual character traits are related to group performance.

Items included in questionnaires of this kind should reflect researchers' conceptualization of the focal collective intellectual character traits. Ideally, the items included in a final questionnaire would be determined through a process that involves drafting a large original pool of items and narrowing this item pool through the use of statistical techniques such as factor analysis. Ideally, the items included in the final questionnaire would exhibit strong properties of reliability, such as a high Cronbach's alpha, high item-scale correlations, and high test-retest correlations. In research of the second kind just described, researchers will want to attend to the extent to which group members converge in sharing a perception of the group. There are common statistical approaches to measuring this kind of inter-rater agreement and reliability, though researchers disagree about whether there is a necessary level of sharedness in perceptions for these perceptions to represent useful data (LeBreton and Senter 2008). In addition to these properties of reliability, researchers would obtain evidence of convergent or divergent validity for the questionnaire by comparing scores on it to scores on other constructs to which they expect it to be similar or different.

With this kind of valid and reliable research instrument available, researchers could then study the relationships between collective

intellectual character traits and other variables of interest, such as outcomes pertaining to group performance or group member experiences. They could attempt to ascertain antecedents of the focal trait and its consequences. They could conduct intervention studies to determine what might affect the presence or absence of the trait in relevant groups. They could conduct longitudinal studies to determine which outcomes are influenced by gains or losses in the trait. Again, this research could study both group members' perceptions of collective intellectual character traits as well as shared perceptions of these traits, where the latter provides a way of measuring the group's own possession of the trait.

Very much this kind of method has been used to study various types of climate in organizations. The study of organizational climate has been defined as the study of "the shared perceptions of and the meaning attached to the policies, practices, and procedures [group members] experience and the behaviors they observe getting rewarded and that are supported and expected" (Schneider et al. 2013, 362). Organizational climate research has been conducted since the 1960s, and two general trends in this research are worth noting here.

First, research on organizational climate has come to emphasize the organizational "level of analysis" rather than the individual level of analysis, which was the focus of some early studies. What this means is that it is attributes of the organization, rather than of individual members in the organization, that are of primary focus in the body of research. Items used in questionnaires focus on attributes of the organization, rather than attributes of the individuals who complete the questionnaires. Indeed, as Schneider and colleagues put it, "Perhaps the major outcome of this area of research for psychology has been the acceptance of a level of theory and data other than the individual as relevant and important in organizational psychological research and practice" (ibid., 369). This focus on the group level of analysis is obviously complementary to the focus on the group level proposed here in the study of collective intellectual character.

Second, research on organizational climate has trended toward the study of "focused climates" rather than "molar climate" (ibid., 365f). Rather than studying organization climate in general, researchers have come to focus on more specific organizational climates that can be connected meaningfully to specific organizational processes or outcomes. Thus, for example, significant scholarly literatures have grown up around safety climate, service climate, diversity climate, and justice climate (Naumann and Bennett 2000). Research has found that these climate features of organizations are indeed related in statistically significant ways to other variables of interest. For example, a higher service climate is predictive of higher customer satisfaction (Schneider et al. 2009), higher safety climate is predictive of fewer accidents and a higher percentage of accidents being reported (Probst et al. 2008), and a more

supportive diversity climate predicts lower gaps in performance between racial/ethnic groups (McKay et al. 2008). These findings not only provide support for the validity of the measures being used to study these particular organizational climates, but they reveal the importance of these organizational climates for the relevant organizations.

Notably, research on organizational climate is not explicitly formulated in terms of organizational character, whether virtuous or vicious or mixed. However, in the growing area of positive organizational scholarship, researchers have attempted to study organizational character explicitly, and they have done so using a methodology very similar to that used in climate research. Kim Cameron is one of the leading researchers in this growing area of research, which he describes as being at its “toddler stage” of development (2017, 430). I’ll describe two illustrative examples of research on collective character that he has conducted in collaboration with others using these methods.

The first example illustrates the first approach identified above, where researchers examine the relationships between individuals’ perceptions of collective character and other variables of interest. Cameron and colleagues take this approach in their (2004), which examines the relationships between perceived organizational virtuousness and organizational performance. Employees from 18 organizations participated in research in which they were asked about the virtuousness of their organizations. They responded to a pool of 60 items created by researchers with expertise in positive organizational scholarship. Researchers used exploratory and confirmatory factor analysis to determine a factorial structure for these items and to create a final, 15-item questionnaire. It contained five subscales representing organizational forgiveness, trust, integrity, optimism, and compassion. Sample items included “Acts of compassion are common here” for compassion, and “Honesty and trustworthiness are hallmarks of this organization” for integrity. Researchers found that perceived virtuousness was a significant predictor of perceived organizational performance, which was itself highly correlated with objective indicators of performance level. In other words, employees who perceived their organizations to be more virtuous also perceived their organizations to have performed better, and the accuracy of their perceptions of organizational performance had independent support.

The second example illustrates the second approach identified above, where researchers use aggregated responses of group members as a measure of collective character traits. Cameron and colleagues take this approach in their (2011), which reports longitudinal studies with financial service units and nursing units focused on the link between virtuousness at the organizational level and organizational effectiveness. Researchers again created a new instrument for measuring organizational virtue, as none had been produced at this time for this kind of study. They drafted an original pool of 114 items assessing what researchers took

to be representative virtuous features of organizations. Sample items included “We treat each other with respect” and “We trust one another”. Using exploratory and confirmatory factor analysis with multiple samples across multiple years, researchers found that the 114 items exhibited a six-factor structure, and could be effectively reduced to a 29-item questionnaire with six group-level character-like subscales: Caring and Kindness; Compassionate Support; Forgiveness; Inspiration and Transcendence; Meaning and Meaningfulness; and Respect, Integrity, and Gratitude. They found that increases in unit scores on these constructs predicted improvements in various dimensions of organizational effectiveness. For example, with nursing units, they predicted improvements in employee turnover, patient satisfaction, employee participation, and quality of care.

These two studies illustrate how the methodology outlined here can be applied to the study of collective character traits, and indeed how it can be done in each of the two ways described above. What I want to propose here is that this same methodology may be fruitfully applied to the study of collective intellectual character traits in particular. It is notable that none of the focal constructs in these two studies is a very good candidate for a collective intellectual virtue. Indeed, I do not know of research in positive organizational scholarship on organizational virtuousness that has had this focus to date. Yet, as the field is still developing, there seems to be a wide-open opportunity for expanding research in the area. And one way this research could be expanded is by incorporating a specific focus on collective intellectual character traits.

We might imagine, for example, research being conducted on the collective intellectual humility of groups of intellectual co-laborers. If we make some simplifying assumptions about similarities between the nature and measurement of collective humility and the nature and measurement of individual intellectual humility, then we might be able to shorten our work somewhat. We can simply adapt existing scales of individual intellectual humility, such as the one discussed earlier, to the collective level. Thus, instead of items such as “When someone points out a mistake in my thinking, I am quick to admit that I was wrong” we would have “When someone points out a mistake in our thinking, we are quick to admit that we were wrong”, and instead of “When I don’t understand something, I try hard to figure it out” we would have “When we don’t understand something, we try hard to figure it out”. If these scales were found to have adequate properties of reliability and validity at the collective level, they could be used to study collective intellectual humility in the way suggested here. In various kinds of studies, researchers could assess the relationships between relevant groups’ intellectual humility and other variables of interest.

A few cautionary notes are, however, in order. And with these, I close.

First, it is possible that, even in cases where we think that a collective intellectual character trait has an analog with an existing measurement instrument at the individual level as in the case with intellectual humility, a simple adaptation of the items from this instrument will not produce a reliable and valid measure of the collective trait. One reason for this is that items that perform well in questionnaires designed for individuals may not perform well when adapted for questionnaires about collectives. For instance, “we” statements have ambiguous readings, and participants may not understand them in the same way or in the way researchers intend. A reference to “our” thinking is ambiguous between referring to how each of us thinks individually versus referring to our shared thinking. Items where this kind of ambiguity is salient, such as “When someone points out a mistake in our thinking, we are quick to admit that we were wrong”, may not perform adequately. Consequently, it may be best to develop new research instruments for collective traits, even if many of the candidate items are closely modeled on items used in existing instruments for individual traits.

Second, it is important to remember our earlier lesson that not all collective intellectual character traits have individual analogs. Thus, for at least some collective intellectual character traits that may be of interest to researchers, it won’t be possible to model items on items contained in questionnaires focused on intellectual character traits of individuals. Questionnaires focused on distinctively collective intellectual virtues may have to be developed *de novo*.

Third and finally, it is important to recall another earlier lesson about different traits being appropriate for different kinds of groups. For example, it would seem that intellectual humility, as conceptualized according to the measure discussed in this chapter, would be a candidate for virtue only for collectives that have learning together as a significant part of their function. This is because intellectual humility, according to this conceptualization, must be motivated by the love of learning. A group that isn’t devoted to learning, even if it is devoted to other intellectual activities such as teaching, may not be a good candidate for being assessed for intellectual humility conceptualized in this way. It may not be an important part of their mission to engage in learning together.

Possibly, we can even learn from cases such as this that different conceptualizations of the character traits we are interested in may be called for. After all, we might think that a group of teachers can display a certain recognizable kind of intellectual humility, even in activities that do not involve them in collaborative learning. We might think they can display intellectual humility in their teaching endeavors, where the relevant kind of humility has to do with a certain kind of service orientation toward learners. If so, we’ll need not just a new measurement instrument

for collective intellectual humility, but a new conceptualization of it—one that is fit for the purpose of studying the particular kind of collective we are interested in.

3 Conclusion

There is a wide-open opportunity to engage in an interdisciplinary study of collective intellectual character. One approach to doing so involves conceptualizing particular collective intellectual character traits, and then operationalizing them and collecting and analyzing data about them using a methodology familiar from research on organizational climate and organizational virtuousness. This methodology mirrors a well-established methodology used to study character traits of individuals. And, this kind of research on collective intellectual character can benefit from consulting related interdisciplinary research on individual intellectual character. Yet, for a variety of reasons outlined in this chapter, it also requires distinctive work of its own. My hope is that this chapter may prompt researchers to experiment with conducting this distinctive work in order to determine whether we may thereby learn more about collective intellectual character.

Notes

- 1 This approach is associated with responsibilist or personalist approaches to virtue epistemology, in contrast to reliabilist approaches (see Battaly 2019), which focus on reliable and unreliable cognitive faculties or belief-producing mechanisms. I am sympathetic with the idea, voiced by Battaly and others, that both reliabilist and responsibilist/personalist approaches have valuable contributions to make to our understanding of excellent (and less than excellent) intellectual functioning. This applies both at the individual level and the collective level. Thus, while my focus here is on collective intellectual character traits, I think there is ample room for valuable contributions to collective epistemology that focus on features other than collective character traits. See chapter also discussions in this volume of research in collective epistemology of a more reliabilist stripe.
- 2 These methods are discussed in more detail in Meyer (2022).

References

- Baehr, Jason. 2011. *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. New York: Oxford University Press.
- Battaly, Heather. 2015. *Virtue*. Malden, MA: Polity Press.
- Battaly, Heather. 2017. "Intellectual Perseverance." *Journal of Moral Philosophy*, 14, 6: 669–697.
- Battaly, Heather. 2019. "A Third Kind of Intellectual Virtue: Personalism." In *The Routledge Handbook of Virtue Epistemology*, ed. Heather Battaly, 115–127. New York: Routledge.

- Bird, Alexander. 2014. "When Is There a Group That Knows? Distributed Cognition, Scientific Knowledge, and the Social Epistemic Subject." In *Essays in Collective Epistemology*, ed. Jenifer Lackey, 42–63. New York: Oxford University Press.
- Byerly, T. Ryan and Meghan Byerly. 2016. "Collective Virtue." *Journal of Value Inquiry* 50, 1: 33–50.
- Byerly, T. Ryan and Meghan Byerly. 2019. "The Collective Characters of Religious Congregations." *Zygon* 54, 3: 680–701.
- Byerly, T. Ryan. 2021. *Intellectual Dependability: A Virtue Theory of the Epistemic and Educational Ideal*. New York: Routledge.
- Cameron, Kim, David Bright, and Aaron Caza. 2004. "Exploring the Relationships between Organizational Virtuousness and Performance." *American Behavioral Scientist* 47, 6: 766–790.
- Cameron, Kim, Carlos Mora, Trevor Leutscher, and Margaret Calarco. 2011. "Exploring the Effects of Positive Practices on Organizational Effectiveness." *Journal of Applied Behavioral Science* 47, 3: 266–308.
- Cameron, Kim. 2017. "Organizational Compassion: Manifestations through Organizations." In *The Oxford Handbook of Compassion Science*, ed. Emma Seppala, Emiliana Simon-Thomas, Stephanie Brown, Daryl Cameron, and James Doty, 421–434. Oxford: Oxford University Press.
- Chan, David. 1998. "Functional Relations among Constructs in the Same Content Domain at Different Levels of Analysis: A Typology of Composition Models." *Journal of Applied Psychology* 83, 2: 234–246.
- Cummings, William and Jung Shin, eds. 2014. "Teaching and Research in Contemporary Higher Education: An Overview." In *Teaching and Research in Contemporary Higher Education: Systems, Activities, Rewards*, ed. Jung Shin, Akira Arimoto, William Cummings, Ulrich Teichler, 1–12. Dordrecht: Springer.
- De Bruin, Boudewijn. 2015. *Ethics and the Global Financial Crisis: Why Incompetence Is Worse than Greed*. Cambridge: Cambridge University Press.
- Furr, Michael. 2011. *Scale Construction and Psychometrics for Social and Personality Psychology*. London: Sage.
- Haggard, Megan, Wade Rowatt, Joseph Leman, Benjamin Meagher, Courtney Moore, Thomas Fergus, Dennis Whitcomb, Heather Battaly, Jason Baehr, and Dan Howard-Snyder. 2018. "Finding Middle Ground between Intellectual Arrogance and Intellectual Servility: Development and Assessment of the Limitations-Owning Intellectual Humility Scale." *Personality and Individual Differences* 124: 184–193.
- Krumrei-Mancuso, Elizabeth, Megan Haggard, Jordan LaBouff, and Wade Rowatt. 2020. "Links between Intellectual Humility and Acquiring Knowledge." *The Journal of Positive Psychology* 15, 2: 155–170.
- Lahroodi, Reza. 2019. "Virtue Epistemology and Collective Epistemology." In *The Routledge Handbook of Virtue Epistemology*, ed. Heather Battaly, 407–419. New York: Routledge Press.
- LeBreton, J. and J. Senter. 2008. "Answers to Twenty Questions about Interrater Reliability and Interrater Agreement." *Organizational Research Methods* 11: 815–852.
- McCrae, Robert and Paul Costa. 1997. "Conceptions and Correlates of Openness to Experience". In *Handbook of Personality Psychology*, ed. R. Hogan, J. Johnson, and S. Briggs, 825–847. San Diego: Academic Press.

- McKay, P., D. Avery, and M. Morris. 2008. "Mean Racial-Ethnic Differences in Employee Sales Performance: The Moderating Role of Diversity Climate." *Personnel Psychology* 61: 349–374.
- Meyer, Marco. 2022. "Measuring Social Virtues." In *Social Virtue Epistemology*, ed. Mark Alfano, Colin Klein, and Jeroen de Ridder. London: Routledge Press.
- Miller, Christian. 2014. *Character and Moral Psychology*. New York: Oxford University Press.
- Naumann, S. and N. Bennett. 2000. "A Case for Procedural Justice Climate: Development and Test of a Multilevel Model." *Academy of Management Journal* 43: 881–889.
- Peterson, Christopher and Martin Seligman. 2004. *Character Strengths and Virtues: A Handbook and Classification*. New York: Oxford University Press.
- Probst, T., T. Brubaker, and A. Barsotti. 2008. "Organizational Under-Reporting of Injury Rates: An Examination of the Moderating Effect of Organizational Safety Climate." *Journal of Applied Psychology* 93: 1147–1154.
- Roberts, Robert and Jay Wood. 2007. *Intellectual Virtues: An Essay in Regulative Epistemology*. New York: Oxford University Press.
- Saucier, Donald and Russell Webster. 2010. "Social Vigilantism: Measuring Individual Differences in Belief Superiority and Resistance to Persuasion." *Personality and Social Psychology Bulletin* 36, 1: 19–32.
- Schneider, B., W. Macey, W. Lee, and S. Young. 2009. "Organizational Service Climate Drivers of the American Customer Satisfaction Index (ACSI) and Financial and Market Performance." *Journal of Service Research* 12: 3–14.
- Schneider, Benjamin, Mark Ehrhart, and William Macey. 2013. "Organizational Climate and Culture." *Annual Reviews in Psychology* 64: 361–388.
- Tanesini, Alessandra. 2018. "Intellectual Servility and Timidity." *Journal of Philosophical Research* 43: 21–41.
- Webster, D. M. and A. Kurglanski. 1994. "Individual Differences in the Need for Cognitive Closure." *Journal of Personality and Social Psychology* 67, 6: 1049–1062.
- Whitcomb, Dennis, Heather Battaly, Jason Baehr, and Dan Howard-Snyder. 2017. "Intellectual Humility: Owning Our Limitations." *Philosophy and Phenomenological Research* 94, 3: 509–539.

15b Commentary from Heather Battaly

Ryan Byerly's chapter maps an interdisciplinary plan for measuring collective intellectual virtues and vices. Byerly argues that measuring such virtues and vices will involve both philosophical work in conceptualizing them, and psychological work—in psychometrics and organizational psychology—in developing measures. Having had a head start on making a philosophical contribution (Byerly and Byerly 2016), Byerly's chapter advocates a conceptualization of collective virtues and vices that is antisummativist, in the sense that it allows a group to have an intellectual virtue or vice that its individual members do not have. With respect to the psychological contribution, Byerly suggests two potential routes for measuring group virtues and vices, one which allows for reports from people outside the group, and another which focuses on the reports of group members. I applaud Byerly's call for future interdisciplinary work on the measurement of collective intellectual virtues and vices. Indeed, it strikes me that the measurement of collective intellectual *vices* is especially needed, as are suggestions about how to ameliorate them! In short, Byerly and I agree about many of the basic parameters. Below I ask three sets of questions about some of his specific proposals.

First, Byerly argues that “what makes a collective intellectual character trait a virtue is that it makes the group that possesses it better as the kind of group that it is”. Likewise, a group's (intellectual) vices will make the group (intellectually) worse as the kind of group that it is. Byerly illustrates this idea by arguing that an intellectual character trait that counts as an intellectual virtue for one group need not count as an intellectual virtue for another group, since the intellectual priorities of groups can differ, for example, in teaching- vs. research-focused universities, and in sketch comedy groups vs. religious congregations. In short, intellectual virtues and vices seem to be indexed to a group's priorities (which identify what kind of group it is). Now, there may be an advantage to this approach when it comes to analyzing liberatory intellectual virtues at the group level. There may also be an advantage to this approach provided that groups prioritize truth, knowledge, understanding, and other epistemic goods. But, what happens when a group's priorities are intellectually bad? When a group's priorities are to hide

or obfuscate the truth, to gaslight marginalized epistemic agents, or to launch campaigns of distraction and misdirection, or to intentionally disseminate falsehoods? Here, troll farms and colluding tobacco executives come to mind. Troll farms seem particularly problematic as their entire existence seems to be predicated on the goal of intentionally disseminating falsehoods. According to the proposal above, dishonesty and epistemic malevolence will count as intellectual virtues for groups with the aforementioned priorities, and honesty and epistemic benevolence will be intellectual vices. But, that seems like the wrong result, or at least an unexpected result since it would make the intellectual virtues and vices of individual agents quite different from the intellectual virtues and vices of groups. Perhaps I have misunderstood Byerly here; in any case, I'd be interested to hear more.

Second, Byerly argues that one reason for endorsing antismmativism is that there are some character traits that are only available to groups and not to individuals, such as character traits that involve regulating and organizing the interactions of group members. Since I am sympathetic with this view, I invite Byerly to expand on the examples he provides, and help us further home in on the features of a trait that would make it exclusively collective. His examples include the intellectual character traits of dividing (well or poorly) intellectual labor among group members, and empowering (or disempowering) group members to contribute to group inquiry. While I wholeheartedly agree that these can count among the intellectual character traits of groups, I wonder whether these particular intellectual character traits are also available to individual leaders within groups. Can't CEOs themselves (and managers at every level) exhibit virtues and vices when it comes to dividing intellectual labor among the members of their departments? Wouldn't this fit de Rooij and de Bruin's call for matching the virtues of individuals with their functions in an organization (2022)? Likewise, can't professors themselves exhibit the virtue (or vice) of empowering (or disempowering) their students to contribute to group inquiry? Perhaps, the point is just that individuals themselves have no members, and thus have no members to divide labor among or empower! Still, individuals can exhibit virtues and vices when it comes to dividing their own intellectual labor (across projects), and empowering themselves to contribute to inquiries (group or otherwise). So, why wouldn't the aforementioned traits of groups be analogs of these? In short, what is it about the aforementioned traits that is supposed to make them only available to groups and not individuals? My hope is that by thinking through some of these examples we can get closer to identifying the features of a trait that make it an exclusively collective trait.

Finally, I close with a set of questions about Byerly's proposals for measuring collective virtues and vices. He argues that one way to do this is to ask individual members of a group to complete surveys about

the group's patterns of action, emotion, perception, and motivation. Their responses could then be aggregated "to determine a score for the group itself", which would serve as a measure of the group's character. For starters, are aggregative methods summative, and if so does that pose a barrier to measuring anti-summative group traits? More importantly, should we be worried about the impact of intellectual vices on the reliability of the responses? If individual group members have intellectual vices, will they be unreliable reporters of the group's actions and motivations? Further, if the group itself has intellectual vices and/or facilitates intellectual vices in its members—if it compartmentalizes information and silos its members, facilitating their ignorance, as star networks do (Sullivan and Alfano 2020), or if it has closed-mindedness and facilitates this vice among its members, as hate groups do—will its members be able to reliably report the group's patterns of action and motivation? Instead of relying on group members to reliably report the group's patterns of action and motivation, should we instead emphasize third-person reports, or coding of the group's actions and motivations? Or, should we be hopeful—will individual group members who are intellectually virtuous be able to detect anti-summative intellectual vices in the group? More broadly, should we, or shouldn't we, expect group vices to be stealthy—to prevent their own detection by the group?

References

- Byerly, Ryan and Meghan Byerly. 2016. "Collective Virtue." *Journal of Value Inquiry* 50: 33–50.
- De Rooij, Barend and Boudewijn de Bruin. 2022. "Real Life Collective Epistemic Virtue and Vice." In M. Alfano, C. Klein, & J. de Ridder (eds.) *Social Virtue Epistemology*. London: Routledge.
- Sullivan, Emily and Mark Alfano. 2020. "Vectors of Epistemic Insecurity." In I. Kidd, H. Battaly, Q. Cassam (eds.) *Vice Epistemology*. Routledge, 148–164.

15c Commentary from Marco Meyer

Measuring Organizational Intellectual Character Traits

Ryan Byerly's chapter makes a proposal for how to empirically study collective intellectual character traits. Part of it is familiar from methodologies to study psychological traits quite generally: Group members are invited to respond to a battery of survey items. For each collective intellectual virtue of interest, the survey contains a few agree/disagree questions. By averaging responses across the respective items, researchers obtain a score per participant for each construct.

Another part of the methodology is specific to capturing *collective* intellectual virtues. Byerly proposes to start from existing instruments that measure individual epistemic virtues, such as humility. Many such instruments use the "direct consensus model", using items written in the first-person singular to encourage respondents to tap answers from their individual perspectives. An example is "When I don't understand something, I try hard to figure it out". If within-group agreement is sufficiently high, scores for the whole group can be obtained by averaging responses across participants.

To capture collective traits, Byerly suggests reformulating items from the perspective of first-person plural: "When we don't understand something, we try hard to figure it out". Some items, Byerly contends, need more adjustment to avoid confusion about what the "we" refers to—each of us individually, or "we collectively". Moreover, some collective traits need to be measured from scratch because there are group virtues such as achieving a good division of intellectual labor that have no equivalent on the individual level.

Is this the right model to study collective intellectual character traits? I will argue that Byerly's proposed methodology is appropriate for the unstructured groups that he mostly discusses, such as groups of educators, religious congregations, or comedy troupes. By contrast, when studying the collective intellectual traits of full-fledged organizations, it is developing the methodology further. The "direct consensus model" that Byerly favors ignores two main distinguishing factors of organizations: hierarchy and division of labor. It thereby risks missing out at the very

features that make organizations intellectually virtuous, or so I shall argue. Adopting the “referent shift consensus model” solves some problems with the “direct consensus model”, but does not address others.

Groups vs. Organizations

What makes intellectual character traits collective? Byerly says that the traits are sensibly attributed to groups. Some groups are unstructured and relatively homogenous. Take a comedy troupe. I picture four people who go on stage together. Byerly suggests that for a comedy troupe, a collective intellectual virtue may be a “witty tendency to detect, appreciate, and satirize each other’s vulnerabilities”. It is characteristic for unstructured and homogenous groups that there is a close link between the perspectives of each group member and the collective intellectual virtues of the group. Looking at average scores on an item like “We like to make fun of others’ vulnerabilities” seems therefore a sensible way of measuring the (self-reported) wittiness of the troupe.

Now suppose the comedy troupe is getting popular. Soon enough it gets requests to perform every evening. Since the comedians are no longer able to manage their bookings and negotiate contracts on the side, they hire strait-laced Tom as manager. I’d like to suggest that the item “We like to make fun of other’s vulnerabilities” would not make much sense to Tom, and not for the reasons that Byerly suggests.

For Tom is not just confused about what “we” refers to. Even if researchers clarified the meaning, neither option—“each of us individually” or “we collectively”—would make sense to Tom. When we consider each member of the troupe individually, there is presumably a huge gulf between the comedians and Tom when it comes to wittiness. Within-group agreement will likely be low because of Tom’s different perspective. Hence, on the direct consensus model, aggregating scores would not be justified, suggesting that there is no collective intellectual virtue of wittiness. Yet this would be a mistake. No doubt, the troupe is as witty as before, and perhaps more so, as the comedians can focus on writing jokes rather than administering bookings.

How would Tom respond if he adopted the second reading of “we”, taking it to refer to the group collectively? Perhaps Tom would agree that the group collectively likes to make fun of each other’s vulnerabilities. Tom does not like to do so, but he may defer to the comedians based on numbers as well as the purpose of the group—it is a comedy troupe, and Tom is the odd one out. But that reaction won’t hold up as the troupe grows in size and complexity. Picture dozens of administrators managing intellectual property, interview slots, and merchandise. Do they, collectively, still like to make fun of each other’s vulnerabilities? Administrators are unlikely to agree, again undermining within-group agreement and thereby aggregation.

The reason that the direct consensus model falters is that in charting the evolution of the comedy troupe, I have transformed the group from an unstructured, homogenous group into an organization, characterized by division of labor and hierarchy. Organizations can be considered as convening spaces for individuals with different epistemic styles. The direct consensus model would perhaps be appropriate if “convening” could be reduced to hiring people with different epistemic styles in the right proportions. But in a well-run comedy troupe, the show does not get less witty every time it hires additional administrative staff. The reason is that there is a division of labor—administrators do not write the jokes or perform on stage. There is also hierarchy. The creative direction of the troupe is set by the comedians. Administrators do not get to vote on jokes.

Shifting Perspective

If the “direct consensus model” is problematic for studying organizations, what are the alternatives? The most prominent contender is the so-called “referent-shift consensus model”. In a first step, items are reformulated to ask respondents to assess the capability of the group: “This troupe produces comedy that makes fun of other people’s perspectives”. This way of formulating items asks respondents to step back from their individual perspective and answer items on behalf of the group, assessing the epistemic traits of the group as if from the outside. As in the “direct consensus model”, in a second step, responses are averaged across respondents, in case there is sufficient within-group agreement.

Adopting the “referent shift consensus model” solves one issue we discussed above: It removes reference to an elusive “we” that assumes that collective intellectual virtues are instantiated by shared values and motivations. Tom may well appreciate that the comedy troupe he works for makes witty comedy. The “referent shift consensus model” allows him to share this perception without asking him to report what he is not: part of a witty “we”.

Yet issues remain with the “referent shift consensus model”. We have shifted the referent, but we have left the consensus approach in place. Hence, we assume that averaging people’s perspectives is a good way to get at the collective intellectual traits of an organization. Whereas I think this approach can be defensible for answering many research questions, we need to be careful for others. People with certain epistemic styles are likely to promote certain collective epistemic traits yet at the same time bias scores in the opposite direction. For instance, I would expect that hiring more lawyers will strengthen intellectual conscientiousness of the organization. Yet lawyers are likely to be more critical on items on items like “This organization carefully considers risks before making a decision”, thereby making the organization seem less cautious.

Generally, adding people with distinct epistemic styles are likely to bias the collective epistemic traits in the organization towards that style, yet are at the same time likely to judge the organization more critically with regard to the presence of their epistemic style than other members.

To overcome the limits of the consensus approach, we could seek to model how epistemic traits in certain subunits of the organization combine to achieve collective epistemic traits at the level of the organization as a whole. This would involve measuring collective epistemic traits at the level of subunits within the organization and developing a theory to draw inferences about the traits of the organization as a whole. There are plenty of subtleties here. For the collective epistemic traits are strongly influenced by seemingly small governance decisions. Are the newly hired lawyers invited to early discussions about new products and partnerships, or are they looped in late in the process? I'd expect that this one decision makes a huge difference for the intellectual traits of the organization. There are likely very many of such levers in organizations, making models of how scores of different subunits interact complicated to operationalize. Because of their simplicity, using the "referent shift consensus model" may be preferable for studying organizations, and the "direct consensus model" for studying unstructured groups. Yet the results should be interpreted with care, and will likely often benefit from accompanying qualitative research.

15d T. Ryan Byerly's Response to Commentaries

I am grateful to Heather Battaly and Marco Meyer for their insightful comments and questions about my chapter. Here I briefly respond to a few of the issues they raise.

First, Battaly raises the question of whether my account of collective virtues implausibly implies that character traits that make bad groups better at being bad are virtues of those groups. I think there are two different replies to this question worth exploring further.

First, it's important to note that my account proposes that collective virtues are traits that make groups better as *the kind* of group they are. Bad groups are typically instances of a salient and more general kind of group which is not bad. For instance, the Nazi party may be a bad group, but the kind of group it is an instance of is a political party, and a political party is not a bad kind of group. Traits that are virtues for the Nazi party, on my view, are not traits that make it better as the Nazi party but traits that make it better as a political party. These won't be traits that make it better at being bad.

Second, it is important to recall Margaret Gilbert and Daniel Pilchman's (2014) lesson that features that are possessed by both individuals and groups may have different qualities—even necessary qualities—in the two cases. These authors caution against arguments that conclude that there can't be collective features of a certain kind because such features couldn't satisfy a necessary condition that applies to them when possessed by individuals. That lesson may apply here too: perhaps at the group level we shouldn't be too quick to grant that virtues (of bad groups) must make their possessors *ceteris paribus* better.

Second, Battaly asks what makes distinctively collective virtues distinctively collective, given that they often do seem to have something like individual analogs. For instance, a group's tendency to distribute intellectual labor well in pursuit of its epistemic aims has an analog in a manager's tendency to guide this very kind of distribution of labor.

In response, I note two points. First, there remain subtle differences between the individual and collective virtues in these cases. The manager's virtue is a tendency to contribute to or guide the distribution of labor among group members in support of the group's intellectual aims, not

a tendency to divide intellectual labor among his/her own members in pursuit of his/her individual intellectual aims. Second, and perhaps more interestingly, the collective virtue seems to have conceptual priority over the relevant individual virtue. To have an account of the manager's virtue of contributing toward excellent division of labor in the group, we first need to know what it is for the group to distribute intellectual labor well. This suggests that in cases of distinctively collective virtues, beginning theorizing with the collective level of analysis is inescapable and will structure theorizing about group member contributions to such collective virtues.

Third, Meyer raises a concern about whether my proposed direct consensus approach to measuring collective virtues may not work when organizations become more complex and distribute labor among more specialized sub-units. For instance, Meyer suggests that a complex comedy organization with specialized sub-units dedicated to managing bookings for the main comedic actors and so forth would remain witty, but that there would be little consensus among group members in this case that the organization as a whole was, say, excellent at spotting and satirizing each other's vulnerabilities. Thus, the direct consensus approach would fail to measure the organization's wittiness.

I think the views of the group members, in this case, are the correct ones. The complex comedy organization does not have the virtue of wittiness. Rather, it has a subtly different virtue that has to do with excellence in *producing* witty comedy, and the direct consensus approach would likely capture this virtue well. It's the main comedic actors alone who retain the group virtue of wittiness. I think this kind of case illustrates the need for interdisciplinarity in the study of group virtues, as without it subtle conceptual distinctions like this can be overlooked easily.

Finally, both Battaly and Meyer raise different kinds of cases where the direct consensus model may falter because of particular details about the kind of trait being measured or the complexion of the group. While I side with Meyer (2022) in thinking that self-report measures, including direct consensus measures, are likely to have widespread utility and are well worth exploring in the first instance due of their simplicity and cost-effectiveness, I also concede that there are cases where these methods are not adequate for measuring group traits.

I'll say a bit more about Meyer's case because I think it illustrates how sometimes only a slight adjustment to the direct consensus approach can overcome the relevant challenge. The case Meyer discusses is one where members of a sub-unit within an organization are high in a particular trait and yet because of this they are more critical of the relative lack of this trait among other group members. Because of this, while the presence of this sub-unit likely increases the group's overall possession of this trait, including these group members' perceptions in a direct consensus approach to measuring the trait may yield lower scores for the group.

I suggest that one approach to overcoming this challenge would involve statistically controlling for it. If group members' own individual traits can be measured alongside their perceptions of the group's trait, and it is known that one variable influences the other, it is easy enough to control for this statistically without pursuing a more complicated approach to measuring group traits. Subtle modifications of the direct consensus approach like this won't always be adequate, but in some cases, they may be.

References

- Gilbert, Margaret and Daniel Pilchman. 2014. "Belief, Acceptance, and What Happens in Groups: Some Methodological Considerations." In J. Lackey (ed.), *Essays in Collective Epistemology*. Oxford: Oxford University Press, 189–213.
- Meyer, Marco. 2022. "Measuring Social Epistemic Virtues: A Field Guide." In M. Alfano, C. Klein, & J. de Ridder (eds.), *Social Virtue Epistemology*. London: Routledge, 523–542.

16 A Bayesian Social Platform for Inclusive and Evidence-Based Decision-Making

S. Kate Devitt, Tamara R. Pearce, Alok Kumar Chowdhury and Kerrie Mengersen

1 Introduction

...when it comes to the direction of human affairs, all these universities, all these nice refined people in their lovely gowns, all this visible body of human knowledge and wisdom, has far less influence upon the conduct of human affairs, than, let us say, an intractable newspaper proprietor, an unscrupulous group of financiers or the leader of a recalcitrant minority.

(Wells 1938)

In January 2021 a mob of supporters of Donald Trump stormed the Capital of the United States (Bergengruen and Time Photo Department 2021). Despite no evidence of electoral fraud, and over 60 failed lawsuits to this effect, the rioters believed that their duty as Americans was to take back their country, to ‘stop the steal’ (Rutenberg et al. 2020; AP/Reuters 2021). The mob believed that Joe Biden had been elected fraudulently, that democracy was at risk and that members of Congress had to be stopped from certifying the electoral votes that would instate Joe Biden as the 46th president of the United States (McSwiney 2021). False beliefs were incubated and amplified not by evidence, but by Donald Trump’s posts on social media platforms, particularly Twitter and Facebook. Once posted on social media, Trump’s messages went viral on social media and via a network of online forums and media creating a ‘right-wing echo chamber’ (Tharoor 2021).

There is no doubt that social media platforms sow disinformation and misinformation just as easily (perhaps much more easily) than true, verifiable information (Singer and Brooking 2018). In the wake of the Capital riots, media commentators have reflected on issues of free speech and moderated content as they pertained to social media (Breton 2021), wondering about the price society pays, particularly democratic societies, when lying becomes normalised (Tenove and McKay 2021).

The unrest in Washington is proof that a powerful yet unregulated digital space—reminiscent of the Wild West—has a profound impact on the very foundations of our modern democracies.

(Breton 2021)

Where years of anguish and lament from ideologues have failed to change misinformation behaviours in the media and social media, corporate litigation has stepped in. Under the threat of defamation lawsuits, media outlets are now changing their behaviours (Brynbaum 2021). Such lawsuits are having an immediate impact on misinformation narratives, for example, during a right-wing media Newsmax interview on 3 February 2021, a host walked off camera to avoid engaging in discussions around unsubstantiated electoral fraud (MSNBC 2021).

Against the backdrop of social media reckoning, this chapter seeks to demonstrate the potential of social tools to build virtuous behaviours online. If we believe that humans would benefit from incorporating philosophical theories into discourse and social knowledge structures, then social media platforms should be created, modified and updated based on our best normative theories in epistemology and the philosophy of science, rather than corporate monetisation metrics. That is to say, the impact of digital content on society should be proportional to the evidence we have for ideas and the comprehensiveness of this evidence. The more justified the ideas (e.g. climate change), the more these ideas should be promoted. If truth matters, then social media platforms must be neither contributor nor content-neutral. And, if they are not neutral, then technology creators and their stakeholders must determine the manner and means of content management.

I take the following ingredients as important to creating good social platforms. First, we must accept humans for the sort of biased actors they are. Humans are myopic, overconfident and affected by contextual factors when they consider ideas (Montibeller and von Winterfeldt 2015). Human behaviour is flawed—and that has to be ok. Second, the truth can be elusive and uncertain, perspectives subjective, evidence contradictory and opinions swayed by *ethos*, *pathos* as well as *logos* (Braet 1992). Third, as communities, we must commit to values, and mechanisms that instantiate these values, to generate an overall society with greater epistemic virtues than our individual behaviours. Societies that use social platforms—either inputting and responding to data or using data produced on them—should value inclusivity, truth (and truth-seeking); and should be receptive to evidence and evidence-based arguments. Platforms must also limit punitive actions and allow productive discord and respectful disagreement.

The ambition to create virtuous social information is not new. The history of information science is largely the instantiation of the dream to collect, collate, store and access the world's best information and documents for the purposes of social good. From oral histories to the written word; shared taxonomies and indexing, to card catalogs and encyclopedias; databases, data mining, business intelligence and expert systems; and more recently recommenders, chatbots and generative language models, humans have sought to store and share good information (Fivush and Haden 2003; Liao 2003; Dacome 2004; Krajewski 2011; Wright 2014; Dale 2021).

Hand-in-hand with virtuous information sharing, particularly since the invention of the printing press, is the parallel spread of propaganda and misinformation through pamphlets, books, newspapers and so forth (Burkhardt 2017). The internet catapulted the potential to deceive and inform, leading scholars to interrogate the factors that need to be considered before one is justified in believing information online (Fallis 2004; Bruce 1997). Information literacy refers to the set of skills and epistemic framework that enable the identification of sources of information; how to access information, and then how to evaluate and use information effectively, efficiently and ethically (Julien and Barker 2009). Information literacy has renewed attention in light of the powerful impacts of ‘fake news’ since 2016 (Jones-Jang et al. 2021; Cooke 2018). New normative frameworks to understand and proactively fight disinformation are emerging (Pamment and Lindwall 2021).

In this chapter, we investigate whether mis- and disinformation can be fought using a social platform that resembles existing platforms, but simultaneously encourages virtuous information behaviours by its design.

The rise of social media in some ways has marked the demise of the document as a primary unit of information (Buckland 1991; Wright 2007). Rather than building up knowledge in expert systems, social media encourages ephemeral, unexpert ejaculations. Social media builds on human gossip mechanisms for shared belief, rather than co-constructing more faithful representations of reality. This chapter suggests a new path for social media in an age of uncertainty and a hunger for evidence-based collective thinking. There is evidence that crowds can be wise, if the circumstances of deliberation and dissent are considered, and mechanisms of groupthink are avoided (Solomon 2006; Sunstein 2011).

Social media success, we argue, is in the hypothesis. The document has long reigned as the unit of information with keywords, indexes and other signals indicating connections to other documents. In the platform we create, the primary unit of information is the hypothesis. Here documents are not intrinsically valuable, but valuable to the degree that they are evidence in service of or to challenge an idea for a purpose (Devitt 2013). Such a reframing allows for and anticipates documents to be error-prone and variable in usefulness in accordance with the ambitions of Bayesian epistemology (Bovens and Hartmann 2004; Hajek and Hartmann 2009; Dunn 2010; Gwin 2011). Centering the hypothesis removes the barrier to using diverse information while limiting the influence of evidence used disproportionately or inappropriately. Traditional social media prioritises the idea too, but to the detriment of evidence and expertise. Social media’s infinite feed of assertions with little evidence creates almost the opposite information environment than that perfected by the book, the document, the card catalogue and the database.

The future of informed conversations requires far-better utilisation of the global ‘world brain’¹ of information through intuitive, yet structured social platforms. To this end, a group of researchers have created a Bayesian social platform for evidence-based collective decision making which we articulate below.

2 Social media

The internet (more broadly) and social media (more specifically) have invited democratic participation in the espousment and evaluation of ideas. Wishing to remain impartial, social media companies have generally welcomed all who wish to register and share their data with them to monetise (Zuboff 2019; Barnet and Bossio 2020). Simple popularity metrics have been employed to adjudicate and share ideas, such as ‘upvoting’ and ‘starring’ content; and retweeting and sharing content within or across platforms. But few features are built or deployed that explicitly work towards improving both the veracity or quality of information shared or the ability of users to effectively evaluate poor information or misinformation. Instead, users share and like information amongst like-minded peers (Schmidt et al. 2017), reducing the friction of dissent and creating epistemic echo chambers. In-group messages expressing righteous or virtuous anger are propagated, while calm, moderate or evidence-based messages are shared less (Singer and Brooking 2018). The science of human behaviours on current dominant social media suggests that, left to their own devices, humans are more likely to reinforce beliefs signalling social group membership/identification and less likely to collectively promote evidence-based beliefs.

This is despite a decade of empirical and theoretical social media research on ways people experience information on platforms such as Twitter and normative guidance for platform producers. For example, Zubiaga and Ji (2014) found that credibility perceptions of tweet authors played a significant role in how users trusted tweets. Basically, the more credible the ‘tweeter’, the more the tweet would be reshared.

Only after 14 years has Twitter added a feature that asks users to employ metacognitive skills, to consider their actions, ‘would you like to read the article before retweeting it?’ In 2020, they ask this question if a user tries to retweet before reading a link (sharing based on trust), rather than opening the link (sharing based on knowledge)—see Figure 16.1.

The experiment with some platforms (starting with Android) went extremely well, with users opening articles 40% more before sharing them, that Twitter has rolled out the feature across all platforms (Hatmaker 2020). Twitter explains this feature because sharing an article can ‘spark conversation’ and opening articles (implied, ‘reading articles’) helps promote informed discussion—see Figure 16.1.



Sharing an article can spark conversation, so you may want to read it before you Tweet it.

To help promote informed discussion, we're testing a new prompt on Android -- when you Retweet an article that you haven't opened on Twitter, we may ask if you'd like to open it first.

4:23 AM · Jun 11, 2020 · Sprinkl

Figure 16.1 Twitter Support tweet explaining the new feature, a prompt to encourage informed discussion. See <https://twitter.com/TwitterSupport/status/1270783537667551233?s=20>.

Social media has traditionally avoided censoring individuals (bad for business) and has allowed networks to grow and their advertising revenue to grow beside it, for example:

At YouTube, we've always had policies that lay out what can and can't be posted. Our policies have no notion of political affiliation or party, and we enforce them consistently regardless of who the uploader is.

(Novacic 2020)

Disregarding political affiliation has led to the rise not only of political extremism but has also made social media the locus of political action such as recruitment, propaganda and collective action. For example, Facebook, Twitter and YouTube were central in the rise of cyber jihadists and Isis (Awan 2017). Facebook enabled warring militias in Libya's civil war to generate and sustain power (Singer and Brooking 2018; Walsh and Suliman 2018). While white supremacists and conspiracy groups such as QAnon in the United States have grown and strengthened with the comprehensiveness of open information on the internet and social media (Hannah 2021). Social media companies do have guidelines to pull down content that includes hate speech, inappropriate content, support of terrorism or spam. But, they also rely on inscrutable decision-making, large cohorts of preciously employed content moderators and automated tools (Gillespie 2018; Roberts 2019; Ganesh and Bright 2020).

However, after the unprecedented mob attack on the US Congress 6 January 2021, incited by weeks of delegitimising the US election,

Twitter first suspended the personal Twitter account of the President of the United States Donald Trump and then permanently deleted it when the user did not obey Twitter's governance rules. Facebook also deleted Trump's accounts and Apple and Google removed the social media app Parlour from its app stores. Amazon removed Parlour from its web hosting services. Within a week of the attacks, thousands of accounts inciting violent insurrection against the US government were removed by Twitter and Facebook.

The question remains whether the solution to social media lies less in content moderation, and perhaps more in the way interaction occurs and information is used. Democratic participation needs to value inclusion and diversity, but also prioritise the knowledge and experience of experts and expertise. Evidence must be drawn from a defensible range of stakeholders and there must be a reasonable opportunity to submit ideas and evidence. Similar to the slow-food movement, future social media must gather and analyse data for propositions over longer temporal periods. The digital social epistemology movement must find a way to encourage interactivity, thoughtfulness and genuine engagement, while also mitigating human cognitive and affect limits, human biases and tendencies.

3 Background

This chapter considers how groups of people might come together more effectively to understand a problem space and to propose actionable solutions from the traditions of social information processing, data-driving decision-making.

3.1 Social information processing

The field of social information processing has long questioned the role of social interaction on social information processing from in-person office interactions to online virtual experiences (Festinger 1954; Salancik and Pfeffer 1978; Meyer 1994; Ahuja and Galvin 2003). Individuals are motivated to communicate with others in order to establish socially derived interpretations for events and their meanings when judgements are important, but the evidence is ambiguous or non-existent and information complex (Salancik and Pfeffer 1978; Meyer 1994). Groups of people desire to fit in and will be motivated to agree with the group. With repetition, ideas are likely to convince individuals, that is make them believe them. Humans use social reasoning as a tool to make sense of uncertainty. Social platforms provide epistemic checking for groups. People will tend to believe what others in their group believe. If evidential reasoning is valued and social reasoning requires evidence, the group may collectively believe propositions for which there is corresponding evidence.

3.2 *Data-driven decisions*

A problem that has arisen across social media and within traditional organisations is that while an overt strategy might recommend ‘data-driven decisions’ (Haller and Satell 2020), in actual fact, decisions are largely made based on political will, trends and biases arising from limited time and resources to evaluate ideas. Even when organisations use data for decisions, often data is incomplete, inaccurate, irrelevant or otherwise problematic to use to base decisions on (Provost and Fawcett 2013). Data is rarely used by itself in raw form, but is transformed via human or machine interpretation, so when we speak of ‘data’ in this chapter, we mean data, models and algorithms; as well as whether data are classed as assertions (aka hypotheses) or evidence for or against hypotheses.

The method specified in this chapter allows groups of humans to use data to make decisions, even when it is partial, messy, and of varying quality. This offers a risk-based approach to data-driven decision-making, where stakeholders are invited to the table but also given a finite timeline. Unlike significance testing in the social sciences, there is no single threshold of evidence under which truth can be presumed; instead, using Bayesian epistemology, beliefs get stronger, the greater the evidence there is to believe in them. Decisions achieve political heft proportionate to the diversity and range of stakeholders invited to contribute and the quantity and quality of contributions.

3.3 *Epistemic justification of social platforms*

How is any information shared on social media justified? Or to put it another way, what gives ideas and information authority, trustworthiness or credibility for decision-makers to progress decisions? Once we can identify what sorts of information we want to see on platforms, then we can consider how to advocate for virtuous online behaviours to manifest better information amongst participants and better management or treatment of this information by decision-makers. This section will go through some of the main sources of justification for information pertinent to digital information sharing.²

For the sake of the chapter, we assume the following:

- P1. **Realism:** basic human beliefs are, for the most part, grounded in perceptions and experiences in the external world that correspond with external reality, for example, humans really see tables, chairs and trees (Devitt 1997; Kornblith 2002) and are not in sceptical conditions (Unger 1978; Audi 2010, Ch. 13–14).
- P2. **Digital Scepticism:** human beliefs are increasingly influenced by veristically-challenged online information environments that

require sceptical vigilance (Cooke 2017, 2018). The saturation of AI-generated (Ippolito et al. 2020), false and misleading digital information increases minimally accurate, inaccurate and false beliefs depending on an agent's ability to curate, manage and correct information flows. Digital scepticism is particularly important information and behaviour promoted by media companies that seek to monetise user attention (Singer and Brooking 2018; Zuboff 2019) and information and behaviours suggested and reinforced by social peers (Eckles et al. 2016; Bailey et al. 2019) and echo chamber effects (Quattrociocchi 2017; Cinelli et al. 2020a, 2020b).

- P3. Justification:** beliefs ought to have both a justified foundation (e.g. via perception, memory, expert testimony) and ought to cohere with other well-justified beliefs (Goldberg 2012; BonJour 2017). Information found in books and online needs to be verified and justified on a case-by-case basis, but influenced by features such as authority, plausibility and support, independent corroboration, and presentation (Fallis 2004, 2006, 2008; Zubiaga and Ji 2014).
- P4. Social Epistemology:** ought to recommend error-correction mechanisms including overriding:
- a singular inaccurate beliefs or poorly grounded beliefs of an individual, for example, where an individual asserts a proposition for which they lack sufficient evidence,
 - b systematic inaccurate or poorly grounded beliefs, for example, an individual or a group of individuals reliably assert propositions with misaligned correspondence with reality or for which they lack sufficient evidence.

Combining these premises, we form a conception of humans interacting in information environments where their connection to reality via traditional modes such as visual perception and memory are grounded by virtue of being evolved to live and succeed in the real world (P1). Yet, human beliefs are increasingly under threat from the deliberate or incidental misinformation from online information environments (P2). In order to be justified in their information habits, humans must develop justified methods to find, sort and evaluate information sourced from a variety of sources (P3). The endeavour to improve epistemic habits is best done within physical and digital social groups (P4). The ambition then is to create a digital infrastructure that provides the sort of justification that holds up to the highest epistemic standards. The benefit of digital tools is that time can be spent honing them over time against our best normative theories.

4 An evidence-based social platform

Researchers set out to make an evidence-based social platform that builds virtuous social information behaviours using interaction mechanisms that instantiate epistemic norms (Devitt et al. 2018). By encouraging social and evidence-based behaviours, the platform sought to build more scientific and inclusive digital cultures. Beginning as a research project, the team was funded by industry and grants to develop a minimal viable product (mvp) and then the minimal marketable product (mmp) for the market, creating a start-up around the platform ‘BetterBeliefs’.³

At its core, BetterBeliefs imagines ideas as hypotheses, represented by horses competing in a ‘hypothesis horse race’. In order to progress in the race, the horses are fuelled by evidence, a little bit like the 20th C. carnival racing game where metal horses compete based on the number of interactions they receive from players (see Figure 16.2.). We thought it would be a breakthrough if data was connected to and presented for or against hypotheses, and data was psychologically engaging, rather than stored in databases hoping for a query to dig it up.

The core functions of the platform for users are:

- Submit hypotheses for consideration
- Submit evidence for and against hypotheses
- Vote on hypotheses to signify approval or disapproval
- Rank the quality of evidence provided for and against hypotheses
- Make a decision based on the degree of belief and weight of evidence of a hypothesis



Figure 16.2 Twentieth century horse race. Image: Casey Hibbard (25 March 2010) <https://www.compelling-cases.com/how-case-studies-get-done-one-leg-at-a-time/>

5 The business case

Organisations ineffectively use the data sets available to them and fail to maximise the value of expensive business intelligence systems (Drucker 1999; Sharma and Djiaw 2011; Richards et al. 2019). While organisations use business intelligence well for budgeting, financial and management reporting, they don't use them for corporate-level decision-making (Richards et al. 2019).

As an Academic start-up dependent on industry funding, the team needed the platform 'to sell', to have a clear value proposition for business. We found evidence that social decision-making and innovation were good for business. For example, crowdsourcing using information systems can support management decision making through several stages of solving a problem (Lindič et al., 2011; Chiu et al. 2014; Ghezzi et al. 2018) such as:

- 1 **Intelligence** (e.g. search, prediction and knowledge accumulation),
- 2 **Design** (e.g. idea generation and co-creation) and
- 3 **Choice** (e.g. voting and idea evaluation) which lead to implementation.

However crowdsourcing can be a double-edged sword, particularly regarding problematic issues such as crowd attitudes and motives; and groupthink and other human biases (Chiu et al., 2014). Crowdsourcing using social platforms may help mitigate some biases in decision-making for innovation but may introduce or exacerbate other biases depending on both platform features and how the platform is used (Bonabeau 2009).

Enterprise Social Media (ESM) is another information system that is a potential mechanism to share ideas across organisational silos, connect people and ideas, and enable innovation. Although the context of ESM is vastly different to commercial social media platforms discussed earlier, the literature on ESM shows that some of the decision-making risk factors for social platforms translate across domains with echo chamber effects and biases including balkanisation and groupthink being highlighted as issues (Leonardi et al. 2013; Leonardi 2014).

The business innovation literature revealed that high ideation rates (having lots of ideas) correlate with growth and net income across organisations. More specifically, there were four key elements essential to high ideation rates (Minor et al. 2017):

- **Scale** (more participants)
- **Frequency** (more ideas)
- **Engagement** (more people evaluating ideas)
- **Diversity** (more kinds of people contributing)

Designing a platform that encouraged these elements of ideation in a social platform that also addressed the thorny issue of effective,

evidence-based decision-making for innovation led to the creation of BetterBeliefs.

6 How does it work?

To design BetterBeliefs, rather than reinventing the wheel of interaction, we selected intuitive mechanisms from existing social media and peer evaluation (e.g. Facebook, Twitter and Reddit). The essential functions of social media are:

- 1 Adding ideas (e.g. text, photos)
- 2 Responding to the ideas of others (e.g. indicating approval by upvoting or clicking on an icon or emoji, replying in a comments section)

On our platform, you can ‘add new hypothesis’ (see Figures 16.3 and 16.4) just like you can create a new tweet on Twitter. But, we built in new significance to the ‘post’ and ‘like’ functions to motivate individuals to think scientifically about claims relevant to their group.

A well-formed hypothesis is a simple proposition that a reasonable person can either agree or disagree with.

E.g. Dogs ought to be the only companion animal allowed on domestic flights inside an aeroplane cabin

When forming hypotheses, we encouraged users to use words that imply what is obligatory, permissible, or forbidden, such as:

Only, most, all, some, many, never, ought, permitted, should, can, needs, should not, cannot, may be, occasionally, sometimes, ought not, in some cases

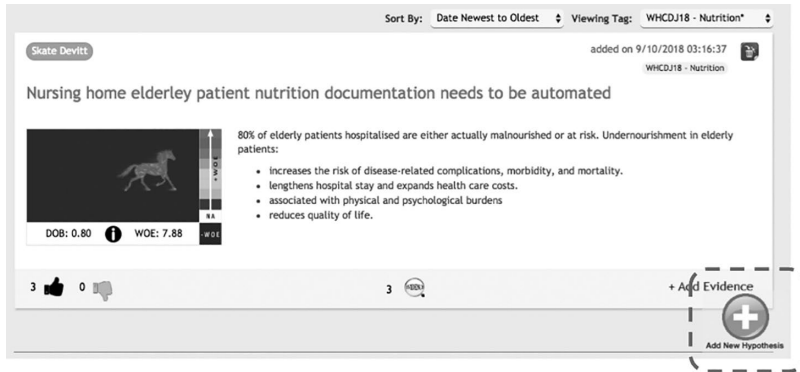
Users add a hypothesis by giving it a title, a tag, some detail and then adding supporting or refuting evidence.

When users add evidence (see Figure 16.5), they provide a URL, a brief argument that explains how their evidence supports or refutes the hypothesis, rank their evidence and identify whether their evidence supports or refutes the hypothesis.

Note encouraging refuting evidence is a key part of BetterBeliefs that we believe no other social platform offers as a mechanism for epistemic evaluation.

Once the platform has hypotheses and evidence, the ‘newsfeed’ view shows users flaming horses and offers an opportunity to ‘thumbs up’ or ‘thumbs down’ the horses.

The degree of belief in the horse is represented by the position of the horse in the black ‘racing box’. A horse to the left-hand side is poorly



Add your ideas to the platform

Figure 16.3 Add a hypothesis to the BetterBeliefs platform.

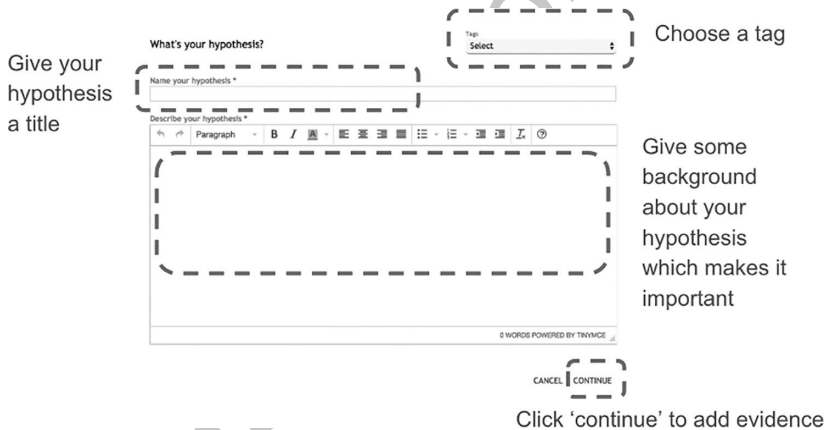


Figure 16.4 Detail your hypothesis.

believed in. A horse to the right-hand side is ‘winning the race’, aka is highly believed in. However, just because a horse is on the right-hand side is insufficient for a win—they need evidence too.

To that end, the horses change colour depending on the weight of evidence for or against them. White horses lack sufficient evidence. Pink horses have much evidence. Blue horses lack evidence. Black horses have evidence largely against them.

For example, a pink horse galloping to the right-hand side of the black box would be a good pick for decision-makers to progress. Whereas a white horse is better ignored until more interactions have occurred on it.

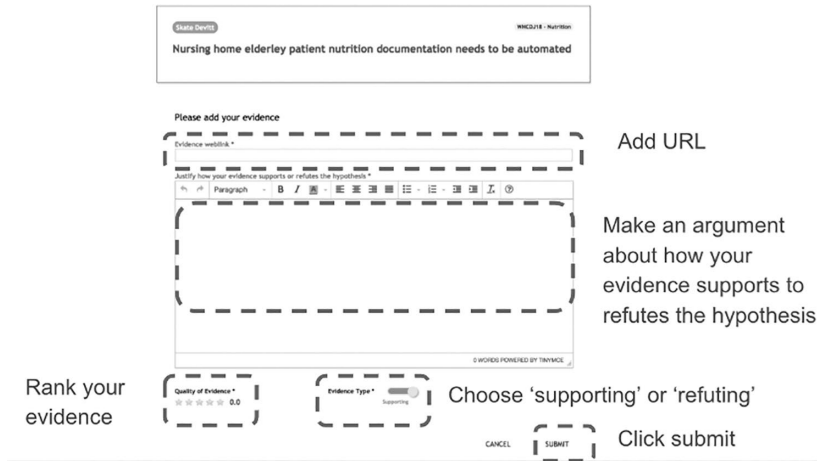


Figure 16.5 Add supporting or refuting evidence.

In fact, a hypothesis will not turn from white to coloured until multiple users have interacted on the hypothesis in terms of both evidence and voting it up or down—see red horse Figure 16.3.

This simple interaction is the basis of the ‘degree of belief’ metric. In aggregate, an organisation or group can understand how much belief there is in a proposition—see Figures 16.6 and 16.7.

The degree of belief (DoB) metric takes the total upvotes and downvotes to create a likelihood that a hypothesis is true given user belief in it using Bernoulli-Beta distributions with 95% credible intervals represented to users. Our confidence in the degree of belief score increases the more users vote hypotheses ‘up’ or ‘down’ (see Figures 16.6 and 16.7).

The sum of evidence (supporting and refuting) plus the quality of evidence added forms the basis of the ‘weight of evidence’ score—see Figure 16.8.

Not all evidence is created equally, so the quality of each piece of evidence must be evaluated to the degree that it supports or refutes hypotheses. When designing the platform, the researchers benefitted from work in statistical science as well as information science on the qualities of information that make it valuable (see Table 16.1). The statistical methods that underpin the platform are commercial in confidence, thus currently not available to the public.

Table 16.1 shows how the team derived six dimensions (credible, accurate, relevant, comprehensive, recent and informative) from Academic research in information systems.⁴ We also created a single star ranking (see Table 16.2) that allowed users to rank evidence based on any

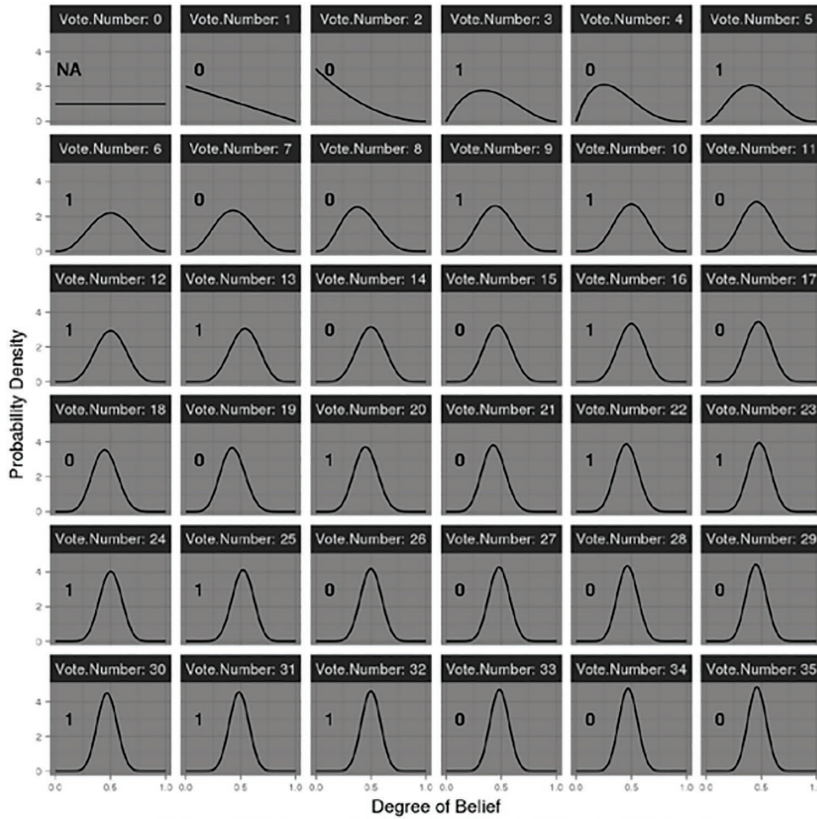


Figure 16.6 How the probability distribution changes for the degree of belief over the first 35 votes on a hypothesis. Credit: Dr Benjamin R. Fitzpatrick.

Table 16.1 Dimensions of information quality and contributing factors for each dimension (Arazy and Kopak 2011; Mai 2013)

<i>Dimension of Information Quality</i>	<i>Contributing Factors for Each Dimension</i>
<i>Credible</i>	Authentic, believable, reliable, trustworthy, authoritative
<i>Accurate</i>	Correct, true, valid
<i>Relevant</i>	Contextual, appropriate
<i>Comprehensive</i>	Complete, objective, neutral, balanced
<i>Recent</i>	Current, up-to-date
<i>Informative</i>	Understandable, useful, usable, good

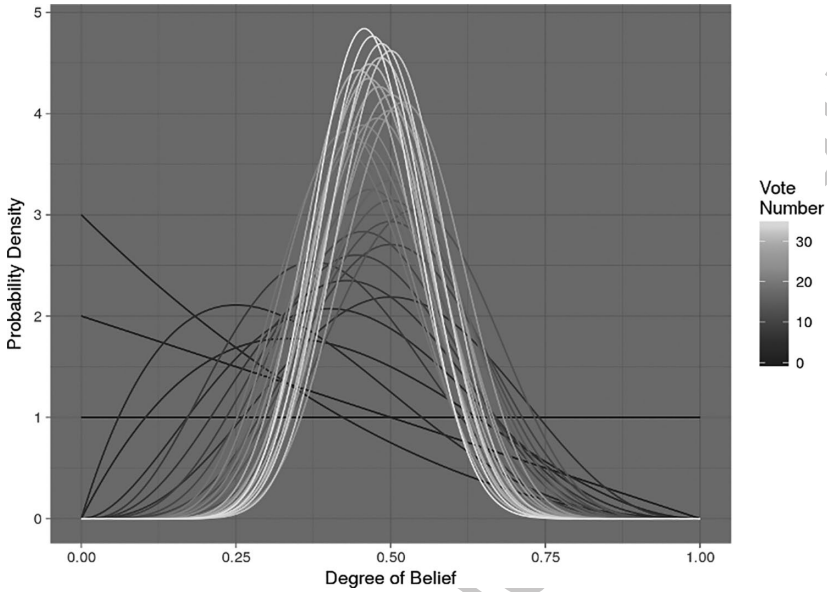


Figure 16.7 The change to probability density as votes are made on hypotheses. Credit: Dr Benjamin R. Fitzpatrick.

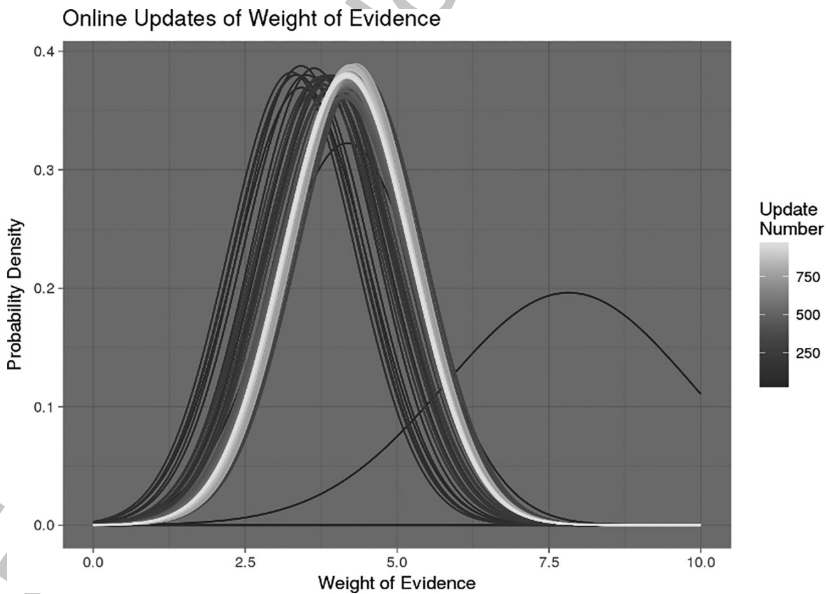










Figure 16.8 The probability density of the weight of evidence as the number of supporting and refuting evidence items of varying quality increases. Credit: Dr Benjamin R. Fitzpatrick. Credit: Dr Benjamin R. Fitzpatrick.

Table 16.2 Guide to ranking evidence items on BetterBeliefs

	Peer reviewed article; Government report
Examples: Australian Journal for Emergency Management; QFES report; Public Choice Journal	
	Industry report
	Investigative or academic journalism; Government media release
Examples: Harvard Business Review; RSPCA	
	Trusted news source
Examples: ABC; BBC; New York Times; Washington Post	
	Generic news source; Substantiated anecdote or case; Expert opinion
Examples: Brisbane Times; Warwick Daily News; Actual example; Leadership anecdotes; Horsetalk.com.au	
	Unsubstantiated anecdote on behalf of community; Well argued point
	Unsubstantiated anecdote on behalf of individual; Argued point
	Opinion
	Feeling

combination of dimensions they felt was relevant to the rank. The theory being that the quality of evidence ranking, in aggregate, would produce ‘better beliefs’ for the collective than either not having the ranking or ranking requiring too much individual effort.

Once users have interacted with both hypotheses and evidence items the Evidence Engine produces the degree of belief (DoB) and weight of evidence (WoE) metrics—see Table 16.3.

Table 16.3 Breakdown of decision quadrants: green, red, amber and white

Category	Description
Green	The green box represents hypotheses that are ‘greenlit for action’ because they meet the decision makers’ threshold for both evidence and belief. Note that the decision maker can use the sliders to change the threshold depending on their own view of what is important for their decision and the consequences for making the decision. If it is a low-risk decision and/or a cheap or easy consequence from the decision, then the decision-maker may set a low threshold. However, if a decision has a lot of risk or the consequences of the decision may involve great costs or time, then the decision maker may require a higher threshold. In each case, due to the inevitable incompleteness of the evidence and limitations of contributors, decision makers will need to satisfice their choice—do ‘enough’ under limitations rather than optimise. They may make threshold decisions based on the number of hypotheses that end up in the green box and/or change the parameters of actions once the decision is made, for example, if all hypotheses are insufficiently evidenced under one reward program, then instead of offering, say seed grants to highly believed hypotheses, they offer a ‘revise-and-resubmit’ to those landing in the green box.
Red	The red box represents hypotheses that are highly believed in yet lack sufficient evidence. A red hypothesis gives the pulse of belief and emotional buy-in. Red hypotheses mean different things depending on the expertise and diversity of participants. If participant intuitions are based on experience, decision makers might divert funds or resources to interrogate why hypotheses are highly believed yet short of evidence. It might be that evidence exists to back up high degrees of belief but has not been added to the platform. Or it might be that beliefs are in fact not sufficiently justified and there is only supposition. Either way decision makers can request users to seek out better evidence for their beliefs or suggest that they downgrade their degree of belief to be commensurate with their evidence.
Amber	The amber box represents hypotheses that have ample evidence, but are not highly believed in. Here, an organisation may wish to <ol style="list-style-type: none"> 1 conduct information or education campaigns to communicate evidence in favour of these beliefs. 2 engage in safe social discussions to combat cognitive dissonance—where individuals are aware of the evidence against their beliefs, but struggle to change them (Beck 2017). 3 encourage unbelievers to add counterevidence to the platform to better justify their beliefs.
White	The white box represents hypotheses that are contentious: have mixed belief or low belief and/or have mixed or limited evidence. There is a diversity of responses to these hypotheses, but the decision maker is unlikely to progress actions on the basis of incomplete or contentious hypotheses. Still, the controversy itself is evidence for decision makers (Christensen 2009). True disagreement offers an opportunity to rethink, reframe and reinvest in seeking good reasons for ideas and taking seriously arguments against them.

The degree of belief is represented between 0.0 and 1.0, where 1.0 indicates 100% belief, absolute certainty in hypothesis 0.5 indicates genuine uncertainty and 0.0 indicates absolute disbelief.

The weight of evidence is on a linear scale with no upper-end limit. This choice is because theoretically there can always be further items of evidence that might increase the likelihood that a hypothesis is true. In reality, users engage with the platform for a finite period of time and there is a limit to the quality and quantity of evidence available to decision makers. Users can view the outputs of the Evidence Engine through the ‘decision dashboard’ that represents hypotheses with increasing degree of belief on the x-axis and increasing weight of evidence on the y-axis. This graphical representation is segmented into quadrants based on thresholds set by the decision maker. The green quadrant has decisions ‘green lit’ for decisions because they meet the threshold for evidence and degree of belief—see Table 16.3.

Finally, users of BetterBeliefs can search the platform for keywords, they can filter hypotheses by recency, degree of belief, the number of evidence items and weight of evidence. Analytics are also available for each hypothesis to get a view of a hypothesis over time.

6.1 Design principles

The platform is designed to:

- 1 Motivate the creation of more relevant options (hypotheses)
- 2 Evaluate options by explicitly linking to evidence
- 3 Harness stakeholder justifications for how the evidence supports or opposes these hypotheses
- 4 Rank evidence to the degree it is (a) quality, (b) relevant to the hypothesis it’s connected with, and c) informative to evaluating hypotheses it is linked with.
- 5 Inform decision-makers about stakeholder ideas and vice versa
- 6 Harness the attraction of social media to teach the scientific method
- 7 Empower groups to make strategic decisions based on stakeholder generated and evaluated hypotheses

6.2 Evidence

A central justification for having beliefs is the degree of evidence one has for them. The more evidence a person has the more they should believe a proposition and vice versa. The greater the risk of having a belief, the more evidence a person should have for that belief. A person should firmly believe a proposition when they have sufficient evidence for it. In general, awareness of one’s evidence for beliefs is considered a good

thing, but the degree to which reflective access is required to be justified is debated (see Dougherty 2011). By focussing on the collective use of evidence, BetterBeliefs treats justification as occurring due to both reliably formed individual beliefs as well as group consideration of explicit evidence and argument.

6.3 *Beliefs*

Traditional epistemology tends to treat beliefs as ‘all-or-none’, either a person believes in p or $\sim p$. Bayesian epistemology takes a different perspective on beliefs. Instead of all-or-none, typical beliefs exist (and are performed) in degrees, rather than absolutes, represented as credence functions. This idea stems from Thomas Bayes who argued that our success in the world depends on how well credence functions, represented in our minds, match the statistical likelihoods in the world (Bovens and Hartmann 2004). This statistical approach to beliefs enables agents to hold multiple beliefs, even contradictory beliefs in their minds at the same time with less certainty. There is evidence that the mind is Bayesian to a certain extent, using adaptive inference to change credence functions in response to evidence (Gopnik and Wellman 2012; Perfors 2012; Clark 2015).

6.4 *Reducing biases*

BetterBeliefs has been designed to reduce cognitive and motivational biases (Kahneman 2011; Montibeller and von Winterfeldt 2015) in hypothesis generation and evaluation by:

- Providing multiple and counter anchors
- Prompting employees to consider reasons in conflict with anchors
- Building explicit probability competence
- Providing counterexamples and statistics
- Capitalising on multiple experts with different points of view about hypotheses
- Challenging probability assessments with counterfactuals
- Probing evidence for alternative hypotheses
- Encouraging decision makers to think about more objectives, new alternatives and other possible states of the future
- Prompting for alternatives including extreme or unusual scenarios

The platform reduces biases algorithmically, for example, differential weightings of users and/or evidence types; interactively, for example, changes to the interface and choice architecture; and culturally, for example, inviting more diverse users or external experts to contribute to the evidence stack and cultural conversations.

Changes to the user interface can reduce biases caused by the way information is displayed and choices are made. Biases can also be reduced culturally through the way the platform is used along with other workshop, ideation and research methods, training events and promotion of virtuous online behaviours by groups. Algorithmic methods to address bias include measurement of user interactions on the system and identifying biased or non-virtuous behaviours.

An example of the algorithmic bias detection potential of the platform is using item-response methods (Embretson and Reise 2013) to identify users that diverge from average response. In an analysis of one-use case of the platform, we could compare the success of ideas posted of sceptical users (those who tended to rate evidence as having less quality than the average user) with ideas posted by generous users (those who tended to rate evidence as having greater quality than average user). Some preliminary, correlative data (protected commercial-in-confidence) suggests that a sceptical culture amongst groups who also engage in prolific hypothesis generation and evaluation may produce more successful ideas than more generous groups.

By encouraging virtuous epistemic behaviours (thinking of many ideas, justifying ideas with evidence and evaluating other people's ideas and evidence) and inhibiting unvirtuous behaviours, the platform ought to reduce a set of biases identified by Montibeller and von Winterfeldt (2015) including: anchoring bias, myopic problem representation, availability bias, omission of important variables, confirmation bias, and overconfidence bias—see Appendix 1. Biases reduced using the Better-Beliefs platform.

Increasing the number and diversity of hypotheses under consideration and encouraging individuals to justify them can improve decision-making even if individual justifications are less than ideal (Oaksford et al. 2016). This comports with a Bayesian approach to evidence, which allows for evidence itself to vary in quality, so long as low-quality evidence is weighted less than higher-quality evidence.

In addition to better hypotheses generation, there are significant benefits to decision makers of having a robust and dynamic set of evaluated hypotheses across teams and work hierarchies to amplify collective intelligence.

6.4.1 Diversity of users

The norms of Bayesian epistemology recommend that more diverse stakeholders and more numerous independent evidential interactions on hypotheses will produce more defensible results to inform decision makers (Bovens and Hartmann 2004; Hajek and Hartmann 2009; Devitt 2013). Diversity of stakeholders can be achieved in three different ways (Pinjani and Palvia 2013):

- 1 demographic or surface-level diversity, for example, age, sex, gender, race,
- 2 deep-level diversity, for example, idiosyncratic attitudes, values and preferences,
- 3 functional diversity, non-overlapping knowledges and expertise in contributors, producing a larger knowledge base on which to draw.

Participants on a successful Bayesian social platform ought to encourage participation from all three kinds of diverse groups, as the likelihood of independence is increased by diversity. Not only did we seek functional diversity, but also to foster the ideas of those on the margins of groups and social networks. Weak ties between individuals have been shown to be good for innovation, whereas strong ties between individuals have been shown to be good for productivity (Granovetter 1973; Levin et al. 2011; Minor et al. 2017).

The platform supposes that the more competent, independent users on the platform considering ideas, the more likely a majority of those users are correct in accordance with Condorcet Jury Theorem (CJT). Condorcet Jury theorem supposes that incorporating the views of many minds (so long as they are competent and independent) will produce truthful propositions.

Not only is diversity important, but so is trust (Palvia, 2009). Contributors must trust that they are able to ‘speak their mind’ and given the benefit of the doubt, be treated with respect, be treated fairly, without unreasonable punitive actions being taken against them.

This method encourages an inclusive, yet evidence-based approach aiming for more reliable and useful results for stakeholders.

6.4.2 *Transparency and access*

Users and decision-makers can download data added to the platform including hypotheses, evidence items, degree of belief, weight of evidence, average quality of evidence, upvotes, downvotes, vote count, rating count, total contributors and authors—see Figure 16.9. Users can choose real names or pseudonyms when they register. The privacy agreement on using the platform models best practice as per GDPR including making the privacy statement as clear as possible.

6.4.3 *Identifying behaviours lacking virtue*

The platform can use algorithmic means to identify online behaviours lacking value, such as:

Careless: a user that endorse hypotheses or pieces of evidence without paying attention

	A	B	C	D	E	F	G	H	I	J	K	L	
1	AddedOn	Title	Description	TagName	DegreeC	Weight	Off	AvgQua	UpVote	DownVi	VoteCo	RatingCo	TotalCi
2	2018-10-09T	Nursing home elderly patient nutrition	80% of elderly patients hospita	Nutrition	0.9	7.6	3.8	5	0	5	4	6	
3	2018-10-09T	Online ads are an effective way to chang	People do change behavior in r	Public Health	0.8	7.6	3.7	2	0	2	3	3	
4	2018-10-09T	We can improve human health & safety	What we eat is really importar	Environment	0.8	8.5	4.0	9	1	10	4	11	
5	2018-10-10T	Public health campaigns should target er	End-of-life care planning for p	Aged Care	0.8	6.8	3.5	4	0	4	2	5	
6	2018-10-11T	Organizations should move to third gene	"In addition to general coping ;	Mental Heal	0.8	6.5	3.3	3	0	3	2	3	
7	2018-10-11T	Investments are needed to prepare fami	Family members are often unq	Home Care	0.8	8.3	4.2	5	1	6	3	6	
8	2018-10-11T	The government should incentivise peopl	As people age, they may maini	Aged Care	0.9	6.4	2.4	7	0	7	2	7	
9	2018-10-11T	Increasing automation in hospital logisti	Can automation help achieve e	Hospital	0.9	7.6	3.8	6	0	6	2	6	
10	2018-10-11T	Reducing the stigma of psychological str	The stigma of psychological sti	Mental Heal	0.8	7.5	3.8	3	0	3	2	3	
11													

Figure 16.9 Sample of downloadable output from the BetterBeliefs platform (authors' names withheld).

Conformity: a user being more likely to upvote hypothesis with high Degree of Belief (DoB) and give a high rank to those with high Weight of Evidence (WoE)

Authorship: a user that downvotes or give low rank to refuting pieces of evidence on a hypothesis they entered and endorsed as well as the inclination to downvote or give low rank to pieces of evidence contrary hypotheses to author's

Group bias and manager fear bias: Users that tend to favour an evidence/hypothesis from their area or added by their direct managers [or anyone higher in hierarchy].

Political coup: a group of individuals acting cooperatively to achieve political ends. This may not be problematic if good and balanced evidence is added. But, detecting such a bias could allow for early intervention on the coup.

Once alerted to poor behaviours, moderators can intervene upon or remove users who are not conforming to community guidelines for online behaviours. There is still much work to be done to ensure moderators have appropriate checks on their own power to influence data production, manipulation and use. Being transparent about how data is generated and used to make decisions is critical in building and maintaining community trust. To date, the BetterBeliefs platform has been used in organisational contexts where corporate, university or government ethics and decision-making is bound by explicit codes of conduct, human resource policy and legislative obligations.

7 Discussion

Virtuous online digital communities seem like a great improvement over apathetic ones, so what could go wrong? In this section, I outline some of the issues that are faced by online content providers and the obligations they have to maintain a just and fair society as well as a knowledge-producing and truth-disseminating one. Key concerns

include the tendency of platforms to exploit user attention and data to progress financial gain (particularly from advertising) to the detriment of user well-being; (Zuboff 2019), the opaque use of surveillance and censorship (Lee and Scott-Baumann 2020), lack of responsibility taken for damaging content posted to and disseminated on platforms in addition to a lack of regulatory oversight. We go through some of these issues in turn.

7.1 Responsibility

Digital platforms have responsibility for both their function and their content. This means that they must have governance structures to evaluate and act on content shared on them if that content is misleading or false as well as causing harm or potentially causing harm. Facebook's Oversight Board is beginning to rule and have impacts on how Facebook manages content, such as the move to remove vaccine misinformation off the platform (Isaac 2021). From responsibility also comes advocacy. Social platforms ought to take a stance on issues (such as public health) and justify behaviours based on this stance. We argue that supporting verifiable content and rejecting demonstrable falsehoods is a critical obligation of social platforms. However, content removal decisions ought to be scrutinised and held to a high standard, lest unwarranted censorship occurs.

7.2 Free speech

Online platforms ought to encourage the free expression of ideas. Mark Zuckerberg has defended the value of free speech to justify *not* taking down posts with problematic content with the exception of posts that could lead to immediate direct physical harm to people on or off the platform. Free speech remains a controversial right as it is frequently misinterpreted as a freedom to say whatever an individual or group wishes to express. On the one hand, freedom is the founding value of the United States where many of the biggest social platforms arose, on the other hand, free speech is misunderstood as including falsehoods and asserting harmful propositions. The Oversight Board has called for Facebook to create more concrete policies that guide their content moderation decisions.

The Board...found Facebook's misinformation and imminent harm rule... to be inappropriately vague and inconsistent with international human rights standards. A patchwork of policies found on different parts of Facebook's website make it difficult for users to understand what content is prohibited.

(Facebook Oversight Board 2021)

Social platforms must abide by the legal obligations in the Sovereign nation within which they are based and abide by International legal frameworks that seek to minimise harms to others. Freedom of expression ought to be endorsed in so far as it maintains authenticity, safety, privacy, dignity and the ability of others to also express themselves.

7.3 Privacy

Online platforms ought to provide privacy to individuals and their content to the degree that users express a preference (Bernal 2014). Such a view would defend a platform for allowing encryption to hide user content as well as allowing users to publicly promote their material. It would also obligate platforms not to conduct unnecessary surveillance or censorship upon users. Platforms must commit to the security of data and information and to resolving data breaches quickly on behalf of users. There are ethical concerns with encryption, such as the wide dissemination with child pornography on communication apps that use encryption. Material that might not be acceptable to the standards of society is likely to be shared via encrypted means. However, encryption also forms a necessary method and means by which citizens can mobilise against an unjust government or fight for their rights as citizens (Daly et al. 2019). Social platforms must remain vigilant with regards to best practice in privacy and security management and vow to continuously update their policies and action to meet the expectations of society and to progress a just and fair society.

7.4 Data rights and data activism

Social platforms ought to be GDPR compliant (or compliant with emerging local governance structures that promote user data rights) (European Parliament and Council 2016). Data subjects ought to be able to request their data and to delete their data. Data activists ought to be able to access and make sense of social platform data creating new ways of knowing the world, creating data countercultures (Milan and Van der Velden 2016). In general, citizens ought to be more empowered to access and use data to progress their ends, particularly the most marginalised and disenfranchised (Daly et al. 2019).

Social platforms can learn from the emerging consensus in ethical AI with regards to how to consider the potential impacts of their technology on the society they serve—see Appendix 2. Comparison of AI Ethics Principles.

To date, the BetterBeliefs platform has been used by organisations for closed groups for specific events including workshops (e.g. Devitt et al. 2021), hackathons, design jams and stakeholder engagement for

strategic policy setting.⁵ In closed settings, moderators and platform designers have worked side-by-side to manage the ethics of platform use and disclosure to users. In the future, the platform team will need to carefully weigh up the excitement of expansion with the ethical risks such an expansion might reveal.

8 Conclusion

Researchers have developed a technology that could be the first step in creating epistemic groups that use social platforms that are inclusive, responsive to evidence, limit punitive actions and allow productive discord and respectful disagreement. BetterBeliefs improves evidence-based, collective ideation—a virtuous digital platform. Our design puts the hypothesis ahead of the document as the unit of information and evidence in the service of or arguing against hypotheses in accordance with the norms of Bayesian epistemology. The platform is designed to help reduce cognitive biases that emerge when groups produce too few hypotheses, hypotheses are too similar or conservative, collective knowledge is ignored, lost or under-utilised, evidence is not comprehensive or is drawn from conforming groups or contexts. Our platform encourages individuals to generate numerous and diverse hypotheses, prompts for different kinds of evidence to support or refute hypotheses, invites users to evaluate the quality of evidence, and scientifically calculates two kinds of metrics for the quality of hypotheses based on how people engage: a ‘degree of belief’ metric that measures how much confidence the group has in a hypothesis; and a ‘weight of evidence’ metric that measures how much evidence the group has considered for or against a hypothesis. The platform can be inclusive, intuitive and rewarding to use. However, while there is potential in using new types of social platforms, platform designs and providers must abide by emerging best practices in social platform governance and responsible innovation, ensuring responsibility, support of free speech, privacy by design, data rights and the opportunity for data activism.

Notes

- 1 See H.G. Wells pre-internet, pre-Wikipedia vision of an updating encyclopedia in every library and institution in ‘World Brain’ (1938).
- 2 Discussions we won’t go into include those around internalism vs. externalism that seek to ground human beliefs against ‘brain in a vat’ style arguments.
- 3 <http://betterbeliefs.com.au>
- 4 During the initial design phase, the team considered inviting users to rate evidence on each dimension, but quickly felt that this would prove too taxing, generating an unwieldy user experience.
- 5 See case studies <https://betterbeliefs.com.au/>

References

- Ahuja, M. K. & Galvin, J. E. 2003. Socialization in virtual groups. *Journal of Management*, 29, 161–185.
- Ap/Reuters. 2021. Facebook to censor ‘stop the steal’ phrase, as social media companies boot US President Donald Trump from their platforms. *ABC News*, 12 January.
- Arazy, O. & Kopak, R. 2011. On the measurability of information quality. *Journal of the Association for Information Science and Technology*, 62, 89–99.
- Audi, R. 2010. *Epistemology: A contemporary introduction to the theory of knowledge*, New York, Routledge.
- Awan, I. 2017. Cyber-extremism: Isis and the power of social media. *Society*, 54, 138–149.
- Bailey, M., Johnston, D. M., Kuchler, T., Stroebel, J. & Wong, A. 2019. Peer effects in product adoption. *National Bureau of Economic Research*, w25843.
- Barnet, B. & Bossio, B. 2020. Netflix’s The Social Dilemma highlights the problem with social media, but what’s the solution? *The Conversation*, 6 October.
- Beck, J. 2017. This article won’t change your mind: The facts on why facts alone can’t fight false beliefs. *The Atlantic*.
- Bergengruen, V. & Time Photo Department 2021. A pro-Trump mob stormed the halls of Congress. Photographs from inside the chaos at the Capitol. *Time*.
- Bernal, P. 2014. *Internet privacy rights: Rights to protect autonomy*, Cambridge, Cambridge University Press.
- Bonabeau, E. 2009. Decisions 2.0: The power of collective intelligence. *MIT Sloan Management Review*, 50, 45–52.
- Bonjour, L. 2017. The dialectic of foundationalism and coherentism. In: Greco, J. & Sosa, E. (eds.), *The Blackwell guide to epistemology*, Chichester, Wiley-Blackwell, 117–142.
- Bovens, L. & Hartmann, S. 2004. *Bayesian epistemology*, Oxford, Oxford University Press.
- Braet, A. C. 1992. Ethos, pathos and logos in Aristotle’s Rhetoric: A re-examination. *Argumentation*, 6, 307–320.
- Breton, T. 2021. Thierry Breton: Capitol Hill—The 9/11 moment of social media. *Politico*, 10 January.
- Bruce, C. 1997. *The seven faces of information literacy*, Adelaide, Auslib Press.
- Brynbaum, M. M. 2021. Lawsuits take the lead in fight against disinformation. *New York Times*, 6 February.
- Buckland, M. K. 1991. Information as thing. *Journal of the American Society for Information Science (1986–1998)*, 42, 351–360.
- Burkhardt, J. M. 2017. History of fake news. *Library Technology Reports*, 53, 5–9.
- Chiu, C.-M., Liang, T.-P. & Turban, E. 2014. What can crowdsourcing do for decision support? *Decision Support Systems*, 65, 40–49.
- Christensen, D. 2009. Disagreement as evidence: The epistemology of controversy. *Philosophy Compass*, 4, 756–767.

- Cinelli, M., Bruognoli, E., Schmidt, A. L., Zollo, F., Quattrociochi, W. & Scala, A. 2020a. Selective exposure shapes the Facebook news diet. *PLoS One*, 15, e0229129.
- Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociochi, W. & Starnini, M. 2020b. Echo chambers on social media: A comparative analysis. *arXiv preprint arXiv:2004.09603*.
- Clark, A. 2015. *Surfing uncertainty : Prediction, action, and the embodied mind*, New York, Oxford University Press.
- Cooke, N. A. 2017. Posttruth, truthiness, and alternative facts: Information behavior and critical information consumption for a new age. *The Library Quarterly*, 87, 211–221.
- Cooke, N. A. 2018. *Fake news and alternative facts: Information literacy in a post-truth era*, Chicago, American Library Association.
- Dacome, L. 2004. Noting the mind: Commonplace books and the pursuit of the self in eighteenth-century Britain. *Journal of the History of Ideas*, 65, 603–625.
- Dale, R. 2021. GPT-3: What's it good for? *Natural Language Engineering*, 27, 113–118.
- Daly, A., Devitt, K. & Mann, M. (eds.) 2019. *Good data*, Amsterdam, Institute for Network Cultures.
- Devitt, M. 1997. *Realism and truth*, Princeton, NJ, Princeton University Press.
- Devitt, S. K. 2013. *Homeostatic epistemology: Reliability, coherence and coordination in a Bayesian virtue epistemology*. Ph.D., Rutgers The State University of New Jersey—New Brunswick.
- Devitt, S. K., Pearce, T. R., Chowdhury, A. & Mengersen, K. 2018. *Strategic decision support platform for collective ideation and evidence-based decisions incorporating Bayesian rationality [decision support tool]*. Institute for Future Environments Queensland University of Technology & The Australian Research Council (ARC), Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS).
- Devitt, S. K., Gan, M., Scholz, J. & Bolia, R. S. 2021. *A Method for Ethical AI in Defence* (DSTG-TR-3786). Available: <https://www.dst.defence.gov.au/publication/ethical-ai>.
- Dougherty, T. 2011. *Evidentialism and its discontents*, Oxford, Oxford University Press.
- Drucker, P. F. 1999. Knowledge-worker productivity: The biggest challenge. *California Management Review*, 41, 79–94.
- Dunn, J. S. 2010. *Bayesian epistemology and having evidence*. Ph.D., University of Massachusetts.
- Eckles, D., Kizilcec, R. F. & Bakshy, E. 2016. Estimating peer effects in networks with peer encouragement designs. *Proceedings of the National Academy of Sciences*, 113, 7316–7322.
- Embretson, S. E. & Reise, S. P. 2013. *Item response theory*, New York, Psychology Press.
- European Parliament And Council. 2016. *General Data Protection Regulation (GDPR)* [Online]. EU. Available: <https://gdpr-info.eu> [Accessed 29 March 2022].
- Facebook Oversight Board. 2021. FB-XWJQBU9A: Case decision 2020-006-FB-FBR. Available: <https://oversightboard.com/decision/FB-XWJQBU9A/> [Accessed 28 January 2022].

- Fallis, D. 2004. On verifying the accuracy of information: Philosophical perspectives. *Library Trends*, 52, 463–487.
- Fallis, D. 2006. Social epistemology and information science. *Annual Review of Information Science and Technology*, 40, 475–519.
- Fallis, D. 2008. Toward an epistemology of Wikipedia. *Journal of the American Society for Information Science and Technology*, 59, 1662–1674.
- Festinger, L. 1954. A theory of social comparison processes. *Human Relations*, 7, 117–140.
- Fivush, R. & Haden, C. A. (eds.) 2003. *Autobiographical memory and the construction of a narrative self: Developmental and cultural perspectives*, Mahwah, NJ, Lawrence Erlbaum.
- Ganesh, B. & Bright, J. 2020. Countering extremists on social media: Challenges for strategic communication and content moderation. *Policy and Internet*, 12, 6–19.
- Ghezzi, A., Gabelloni, D., Martini, A. & Natalicchio, A. 2018. Crowdsourcing: A review and suggestions for future research. *International Journal of Management Reviews*, 20, 343–363.
- Gillespie, T. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*, New Haven, CT, Yale University Press.
- Goldberg, S. 2012. A reliabilist foundationalist coherentism. *Erkenntnis*, 77, 187–196.
- Gopnik, A. & Wellman, H. M. 2012. Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, 138, 1085–1108.
- Granovetter, M. S. 1973. The strength of weak ties. *American Journal of Sociology*, 78, 1360–1380.
- Gwin, M. F. 2011. *The virtues of Bayesian epistemology*. Ph.D., University of Oklahoma.
- Hajek, A. & Hartmann, S. 2010. Bayesian epistemology. In: Dancy, J., Sosa, E. & Steup, M. (eds.) *A companion to epistemology*. Chicester, John Wiley & Sons, 93–105.
- Haller, E. & Satell, G. 2020. Data-driven decisions start with these 4 questions. *Harvard Business Review*, 11 February.
- Hannah, M. 2021. QAnon and the information dark age. *First Monday*, 26.
- Hatmaker, T. 2020. Twitter plans to bring prompts to ‘read before you retweet’ to all users. *Techcrunch*, 25 September.
- Ippolito, D., Duckworth, D., Callison-Burch, C. & Eck, D. 2020. Automatic detection of generated text is easiest when humans are fooled. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 1808–1822.
- Isaac, M. 2021. Facebook says it plans to remove posts with false vaccine claims. *New York Times*, 8 February.
- Jobin, A., Ienca, M. & Vayena, E. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399.
- Jones-Jang, S. M., Mortensen, T. & Liu, J. 2021. Does media literacy help identification of fake news? Information literacy helps, but other literacies don’t. *American Behavioral Scientist*, 65, 371–388.

- Julien, H. & Barker, S. 2009. How high-school students find and evaluate scientific information: A basis for information literacy skills development. *Library & Information Science Research*, 31, 12–17.
- Kahneman, D. 2011. *Thinking, fast and slow*, New York, Farrar, Straus and Giroux.
- Kornblith, H. 2002. *Knowledge and its place in nature*, New York, Oxford University Press.
- Krajewski, M. 2011. *Paper machines: About cards & catalogs, 1548–1929*, Cambridge, MA, MIT Press.
- Lee, Y. & Scott-Baumann, A. 2020. Digital ecology of free speech: Authenticity, identity, and self-censorship. In: Yates, S. J. & Rice, R. E. (eds.) *The Oxford handbook of digital technology and society*, New York, Oxford University Press, 470–497.
- Leonardi, P. M. 2014. Social media, knowledge sharing, and innovation: Toward a theory of communication visibility. *Information Systems Research*, 25, 796–816.
- Leonardi, P. M., Huysman, M. & Steinfield, C. 2013. Enterprise social media: Definition, history, and prospects for the study of social technologies in organizations. *Journal of Computer-Mediated Communication*, 19, 1–19.
- Levin, D. Z., Walter, J. & Murnighan, J. K. 2011. Dormant ties: The value of reconnecting. *Organization Science*, 22, 923–939.
- Liao, S.-H. 2003. Knowledge management technologies and applications—Literature review from 1995 to 2002. *Expert Systems with Applications*, 25, 155–164.
- Lindič, J., Baloh, P., Ribière, V. M. & Desouza, K. C. 2011. Deploying information technologies for organizational innovation: Lessons from case studies. *International Journal of Information Management*, 31, 183–188.
- Mai, J. E. 2013. The quality and qualities of information. *Journal of the Association for Information Science and Technology*, 64, 675–688.
- Mcswiney, J. 2021. Why were the Capital rioters so angry? Because they're scared of losing grip on their perverse idea of democracy. *The Conversation*, 8 January.
- Meyer, G. W. 1994. Social information processing and social networks: A test of social influence mechanisms. *Human Relations*, 47, 1013–1047.
- Milan, S. & Van Der Velden, L. 2016. The alternative epistemologies of data activism. *Digital Culture & Society*, 2, 57–74.
- Minor, D., Brook, P. & Bernoff, J. 2017. Data from 3.5 million employees shows how innovation really works. *Harvard Business Review*.
- Montibeller, G. & Von Winterfeldt, D. 2015. Cognitive and motivational biases in decision and risk analysis. *Risk Analysis*, 35, 1230–1251.
- Msnbc. 2021. *MyPillow fight: Lindell clashes with newsmax over Trump's 2020 loss | The 11th hour | MSNBC* [Online]. YouTube: YouTube. Available: <https://youtu.be/jQNVSitbrY4> [Accessed 8 February 2021].
- Novacic, I. 2020. Censorship on social media? It's not what you think. *CBS News*, 28 August.
- Oaksford, M., Singmann, H., Douven, I., Krzyzanowska, K., Hahn, U. & Lombrozo, T. 2016. Dynamic inference and belief revision. *International Conference on Thinking*, 4–6 August 2016. Brown University.

- Palvia, P. 2009. The role of trust in e-commerce relational exchange: A unified model. *Information & Management*, 46, 213–220.
- Pamment, J. & Lindwall, A. K. 2021. *Fact-checking and debunking: A best practice guide to dealing with disinformation*. NATO Strategic Communications Centre of Excellence.
- Perfors, A. 2012. Bayesian models of cognition: What's built in after all? *Philosophy Compass*, 7, 127–138.
- Pinjani, P. & Palvia, P. 2013. Trust and knowledge sharing in diverse global virtual teams. *Information & Management*, 50, 144–153.
- Provost, F. & Fawcett, T. 2013. Data science and its relationship to big data and data-driven decision making. *Big Data*, 1, 51–59.
- Quattrociocchi, W. 2017. Inside the echo chamber. *Scientific American*, 316, 60–63.
- Richards, G., Yeoh, W., Chong, A. Y. L. & Popovič, A. 2019. Business intelligence effectiveness and corporate performance management: An empirical analysis. *Journal of Computer Information Systems*, 59, 188–196.
- Roberts, S. T. 2019. *Behind the screen: Content moderation in the shadows of social media*, New Haven, CT, Yale University Press.
- Rutenberg, J., Corasaniti, N. & Feuer, A. 2020. Trump's fraud claims died in court, but the myth of stolen elections lives on. *New York Times*, 26 December.
- Salancik, G. R. & Pfeffer, J. 1978. A social information processing approach to job attitudes and task design. *Administrative Science Quarterly* 23(2), 224–253.
- Schmidt, A. L., Zollo, F., Del Vicario, M., Bessi, A., Scala, A., Caldarelli, G., Stanley, H. E. & Quattrociocchi, W. 2017. Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences*, 114, 3035–3039.
- Sharma, R. S. & Djiaw, V. 2011. Realising the strategic impact of business intelligence tools. *VINE Journal of Information and Knowledge Management Systems*, 41, 113–131.
- Singer, P. W. & Brooking, E. T. 2018. *LikeWar: The weaponization of social media*, New York, Houghton Mifflin Harcourt.
- Solomon, M. 2006. Groupthink versus the wisdom of crowds: The social epistemology of deliberation and dissent. *The Southern Journal of Philosophy*, 44, 28–42.
- Sunstein, C. R. 2011. Deliberating groups versus prediction markets (or Hayek's challenge to Habermas). In: GOLDMAN, A. & WHITCOMB, D. (eds.) *Social epistemology: Essential readings*. New York, Oxford University Press, 314–337.
- Tenove, C. & Mckay, S. 2021. Trump's lies about the election show how disinformation erodes democracy. *The Conversation*, 30 November.
- Tharoor, I. 2021. The need to reckon with Trump's lies. *The Washington Post*, 11 January.
- Unger, P. 1978. *Ignorance: A case for scepticism*, Oxford, Oxford University Press.
- Walsh, D. & Suliman, A. Z. 2018. A Facebook war: Libyans battle on the streets and on screens. *New York Times*, 4 September.
- Wells, H. G. 1938. *World brain*, London, Methuen & Co., Ltd.

Wright, A. 2007. *Glut: Mastering information through the ages*, Washington, DC, Joseph Henry Press.

Wright, A. 2014. *Cataloging the world: Paul Otlet and the birth of the information age*, Oxford/New York, Oxford University Press.

Zubiaga, A. & Ji, H. 2014. Tweet, but verify: Epistemic study of information verification on twitter. *Social Network Analysis and Mining*, 4, 163.

Zuboff, S. 2019. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*, New York, PublicAffairs.

T&F Proofs – Not for Distribution

Appendix 1

Biases Reduced Using the BetterBeliefs Platform

Anchoring bias occurs when the estimation of a numerical value is based on an initial value (anchor), which is then insufficiently adjusted to provide the final answer.

Found in estimation tasks, pricing decisions and negotiations.

Ways to debias: avoiding anchors, providing multiple and counter anchors, and using experts with different anchors. Prompt employees to identify features of the target variable different than the anchor, or to consider reasons in conflict with the anchor.

Availability bias (or 'ease-of-recall') occurs when ease of recall dominates the assignment of probability to an event.

Found in frequency estimates, frequency of lethal events and rare events that are anchored on recent examples.

Ways to debias include conducting probability training, providing counterexamples and providing statistics.

Confirmation bias occurs when there is a desire to confirm one's belief by selectively acquiring and using evidence.

Found in many settings such as information gathering, selection tasks, evidence updating and evaluation of one's own judgement. It has been shown in real-world contexts such as medical diagnostics, judicial reasoning and scientific thinking.

Ways to debias confirmation bias include using multiple experts with different points of view about hypotheses, challenging probability assessments with counterfactuals and probing evidence for alternative hypotheses.

Myopic problem representation occurs when an incomplete mental model creates an oversimplified problem representation.

Found when participants focus on a small number of alternatives, a small number of objectives or a single future state of the world.

Ways to debias trying to encourage decision makers to think about more objectives, new alternatives and other possible states of the future.

Omission of important variables occurs when an important variable is overlooked.

Found in the definition of objectives, identification of decision alternatives and hypothesis generation. Ways to debias prompt for alternatives and objectives, ask for extreme or unusual scenarios or use group elicitation techniques.

Overconfidence bias occurs when the decision makers provide estimates for a given parameter that are above the actual performance (overestimation) or when the range of variation they provide is too narrow (over precision).

Found frequently in quantitative estimates, such as in defence, legal, financial and engineering decisions. Also present in judgements about the completeness of a hypothesis set.

T&F Proofs – Not for Distribution

Appendix 2

Comparison of AI Ethics Principles

<i>Australian Government's AI Ethics Principles (Department of Industry Innovation and Science 2019)</i>	<i>Principled Artificial Intelligence: A Map of Ethical and Rights-Based Approaches</i>	<i>The Global Landscape of AI Ethics Guidelines (Jobin et al. 2019)</i>
Human, social and environmental wellbeing: Throughout their lifecycle, AI systems should benefit individuals, society and the environment Human-centred values: Throughout their lifecycle, AI systems should respect human rights, diversity, and the autonomy of individuals	Promotion of human values Professional responsibility	Responsibility
Transparency and explainability: There should be transparency and responsible disclosure to ensure people know when they are being significantly impacted by an AI system, and can find out when an AI system is engaging with them	Human Control of Technology Transparency	Transparency
Reliability and safety: Throughout their lifecycle, AI systems should reliably operate in accordance with their intended purpose	Fairness and non-discrimination Safety and Security	Justice and fairness Non-maleficence
Fairness: Throughout their lifecycle, AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups Privacy protection and security: Throughout their lifecycle, AI systems should respect and uphold privacy rights and data protection, and ensure the security of data Contestability: When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or output of the AI system	Privacy	Privacy
Accountability: Those responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled	Accountability Explainability	

16b Commentary from Jeroen de Ridder

BetterBeliefs and Cognitive Hooligans

The work that Kate Devitt and her colleagues report in their chapter is absolutely fascinating. Anyone in their right mind who has spent time on Twitter or other social media platforms will share a heartfelt wish for these platforms to do better at promoting truth-seeking, rather than misinformation cascades or polarising interaction. This is exactly the vista of a better world that BetterBeliefs promises. Since Bayesianism forms BetterBeliefs' beating heart, I wouldn't expect problems to arise there. I want to think about how people might use a platform like this or, more specifically, *which* people are most likely to use it and how that might pose problems.

It's one of the most consistent and widely confirmed findings in political science over the past half-century that most people are largely ignorant about politics (Somin 2016; Achen and Bartels 2017). But this sad state of affairs doesn't prevent *some* people from being highly informed about politics. For the most part, they are people who find politics interesting and entertaining. They seek out lots of political information to stay abreast of current political affairs, they enjoy discussing politics with their friends, and they derive satisfaction from making sense of what happens in politics.

So far, so good. Unfortunately, it turns out that people aren't very good at processing political information in a reliable—that is, unbiased, fair-minded, and objective—manner. On the contrary, people tend to be *tribalistic* cognisers; they process information in highly biased ways. Social identity and partisan alliance come first, truth second. This has been shown over and over again in psychology and political science. The picture emerging from this literature is that people's engagement with politics is more like that of sports fans or religious devotees than that of scientists or investigators. The political philosopher Jason Brennan's take on the evidence is that many 'political fans' resemble *hooligans*—and he is hardly unique (cf. Mutz 2006; Caplan 2011; Somin 2016; Achen and Bartels 2017):

They have strong and largely fixed worldviews. They can present arguments for their beliefs, but they cannot explain alternative points

of view in a way that people with other views would find satisfactory. Hooligans consume political information, although in a biased way. They tend to seek out information that confirms their preexisting political opinions, but ignore, evade, and reject out of hand evidence that contradicts or disconfirms their preexisting opinions. (Brennan 2017, 5)

Contrary to what one might expect—or at least hope—these tendencies are most pronounced among the most knowledgeable: ‘the most knowledgeable voters tend to be more biased in their evaluation of new evidence than those with less prior political information’ (Somin 2016, 80; cf. Taber and Lodge 2006) and ‘those most knowledgeable about and interested in politics are not the people most exposed to oppositional political viewpoints. The dominance of like-minded over oppositional voices increases as political knowledge increases’ (Mutz 2006, 32).

Thinking about politics isn’t all that different from thinking about other things that interest us and that might matter to our social identity. There is plenty of evidence that myside bias and tribalistic information processing are rampant in many areas of our cognition (Haidt 2012; Mercier and Sperber 2017; Slovic and Fernbach 2017; Stanovich 2021).

If this is right, it spells trouble for BetterBeliefs. First, the kinds of things it does and how it does them will make it (more) attractive to certain kinds of people, to wit people who enjoy analysing problems, gathering, sharing, and weighing evidence pro and con, and discussing these things with others. In other words, people who tend to be well-informed (about the things they take an interest in) and enjoy analytic-reflective thinking. This can affect the hypotheses and evidence that end up being considered on BetterBeliefs. In politics, it turns out that being more informed has systematic effects on one’s preferences and beliefs (Althaus 2003; Bovens and Wille 2017). In so far as this generalises to other domains, there is a real risk that BetterBeliefs users have systematically different interests, beliefs and preferences than non-users, which will skew its outputs.

Second, high-information reflective users might not just start out with one-sided and biased beliefs, but—as we saw above—are also prone to process new information in biased ways, playing up evidence in favour of their prior beliefs and downplaying evidence speaking against it. We should expect this to characterise their usage of BetterBeliefs, too: entering hypotheses and evidence that fit prior beliefs and interpreting and weighing evidence in light of these prior beliefs. Since the platform depends on its users’ input of hypotheses, evidence, and assessments of their quality and credibility, it might end up reinforcing biased modes of cognition.

This suggests that whether BetterBeliefs can ultimately do an epistemically better job than extant social media platforms will depend crucially

on at least two things. First, on whether it can attract a sufficiently diverse user base—including, importantly, users who may not be naturally given to the sort of evidence-based, reflective cognition BetterBeliefs seeks to support. And second, on whether its anti-biasing measures will prove sufficiently powerful to offset the cognitive hooliganism that high-information individuals tend to engage in.

References

- Achen, Christopher H., and Larry M. Bartels. 2017. *Democracy for Realists*. Princeton, NJ: Princeton University Press.
- Althaus, Scott L. 2003. *Collective Preferences in Democratic Politics*. Cambridge: Cambridge University Press.
- Bovens, Mark A.P., and Anchrit C. Wille. 2017. *Diploma Democracy: The Rise of Political Meritocracy*. New York: Oxford University Press.
- Brennan, Jason. 2017. *Against Democracy*. Princeton, NJ: Princeton University Press.
- Caplan, Bryan. 2011. *The Myth of the Rational Voter*. Princeton, NJ: Princeton University Press.
- Haidt, Jonathan. 2012. *The Righteous Mind*. New York: Pantheon Books.
- Mercier, Hugo, and Dan Sperber. 2017. *The Enigma of Reason*. Cambridge, MA: Harvard University Press.
- Mutz, Diana C. 2006. *Hearing the Other Side*. Cambridge: Cambridge University Press.
- Slooman, Steven, and Philip Fernbach. 2017. *The Knowledge Illusion: Why We Never Think Alone*. New York: Riverhead.
- Somin, Ilya. 2016. *Democracy and Political Ignorance*. 2nd ed. Stanford, CA: Stanford Law Books.
- Stanovich, Keith E. 2021. *The Bias That Divides Us*. Cambridge, MA: MIT Press.
- Taber, Charles S., and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50 (3): 755–769. <https://doi.org/10.1111/j.1540-5907.2006.00214.x>.

16c Commentary from Erik J. Olsson

In their fascinating and inspiring contribution to this volume, Kate Devitt and her colleagues Tamara Pearce, Alok Chowdhury and Kerrie Mengersen describe a Bayesian social platform—BetterBeliefs—for inclusive and evidence-based decision-making. The implementation features a Twitter-like interface, although there are several key points where it differs from Twitter in its functionality. The platform is perhaps not so much an academic achievement as it is an attempt to bring social epistemology to the world of organisations and enterprises, a most laudable goal as I see it.

A tool suitable for an enterprise needs to be useful in a way that connects with the goals of the enterprise. BetterBeliefs implements this practical requirement by focusing on what Devitt et al. call ‘hypotheses’. The term hypothesis here denotes something like a policy proposal rather than a factual claim. To take one of the authors’ own examples, ‘Nursing home elderly patient nutrition documentations needs to be automated’ would count as a hypothesis. A user may add such a hypothesis in a box in the interface, together with some background information about why it is important. The next step is to add evidence vis-à-vis the hypothesis. This is accomplished, first, by adding a URL to an online document in which evidence is to be found and, second, by providing an argument in a textbox as to how the evidence connects to the hypothesis. The user is then asked, if I understand it correctly, to rank the strength of the evidence (on a scale from ‘feeling’ to ‘peer-reviewed article: government report’) and to indicate whether the evidence is supporting or refuting. Other users can provide evidence for or against a given hypothesis as well as upvote or downvote it, illustrating the ‘inclusive’ aspects of the platform. An organisational tool stands a greater chance of success if it is easy and fun to work with. BetterBeliefs represents the current credibility status of hypotheses using the amusing analogy with a horse race, each horse representing a hypothesis. The colour of a horse depends on the votes and evidence provided for or against the hypothesis that it represents.

From a research perspective, the way in which the votes and the evidence are used is perhaps the most interesting feature of the platform. There is

an in-built function taking the total number of upvotes and downvotes as input to compute the likelihood that a hypothesis is true (using a Bernoulli-Beta distribution to represent the current uncertainty in the degree of belief). The more votes are in, the greater the confidence in the degree of belief resulting from the votes. Users' indications of the strength of evidence provide input to a Bayesian machinery that computes the total strength of evidence for the hypothesis. There are some philosophical issues here regarding how and in what sense a policy proposal can be true or false and, on that basis, assigned a 'likelihood', but there are other things that I would rather discuss in this brief commentary, so I move on.

At this point, it would of course have been interesting to see the underlying equations for computing the overall strength of the evidence. However, we are told that '[t]he statistical methods that underpin the platform are currently not available to the public'. The openness that we take for granted in academic research no longer holds in the business world where innovations are carefully guarded trade secrets. The effect is that it is not possible to evaluate the reasonableness of the statistical methods used from a social-epistemological perspective. All we know is that the methods are broadly Bayesian, a statement that is compatible with a whole range of approaches, some more plausible or rational than others.

Yet, I think it would be a failure of judgement to hold the secrecy issue against the authors. Rather, they are playing a different game than we are used to for which a different set of rules apply. A further difference to academic work, besides secrecy, is that for business purposes a platform does not have to be perfect in every detail, so long as it is better than what is already out there. My knowledge of the business world is too limited for me to be in a position to assess whether BetterBeliefs beats its competitors, if there are any. What I can say is that I think Devitt and her colleagues have done a remarkable job in addressing the difficult challenges and design choices that inevitably arise when translating academic research into a workable product.

I am slightly less convinced by the way in which Devitt et al. choose to frame their case for BetterBeliefs from a social perspective. First, some of their more general remarks about the failure of social media to prevent misinformation seem rather far-reaching but more importantly not always clearly connected to their own platform and its actual functionality. As Devitt et al. describe BetterBeliefs, any hypothesis can be introduced and voted on, and any evidence for or against it can subsequently be added. Furthermore, any person can, in principle, contribute by introducing new hypotheses, or react to hypotheses already introduced. Moreover, the purpose of the platform is to provide 'better beliefs' in the sense of beliefs that are better adjusted to the actual evidence and therefore more 'truthful', to use the authors' own term. To my mind, however, all this stands in contrast to the authors' statement in the introduction that '[i]f truth matters, then social media platforms must be

neither contributor nor content neutral'. It seems to me that BetterBeliefs is precisely a contributor and content-neutral platform for people and organisations for whom truth matters. It is correct, though, that the authors, at the end of their paper, discuss various strategies for moderation at the contributor level, for example removing users who are not conforming to 'community guidelines for online behaviors', but it is unclear whether the information that is extractable about user behaviour in BetterBeliefs can credibly inform such action.

Second, while much of what the authors write on more general issues such as free speech, responsibility, diversity and so on may sound good and laudable at first sight, the devil is surely in the details. For example, the authors note that freedom of speech is a founding value of the United States where many of the biggest social platforms arose, and yet, they claim, 'free speech is misunderstood as including falsehoods and asserting harmful propositions'. However, free speech does in fact not prohibit asserting falsehood or harmful propositions *per se*. A general prohibition against asserting falsehood or harmful propositions would have a chilling effect on free speech in violation of, for example, Article 10 of the European Convention of Human Rights. Only in particular cases, which in the legal frameworks of modern liberal democracies need to be described in sufficiently precise terms, is false or harmful speech prohibited, for example in the interest of protecting a citizen's reputation, in which case a precise law to this effect needs to be in place.

Third, while I realise that a reference to Donald Trump and the storming of the Congress are obligatory ingredients in any account of the ills of fake news etcetera, I would have welcomed examples of irresponsible online behaviour coming from different areas of the political spectrum. There is no shortage of examples from leftwing groups regarding, for example genetically modified organisms (GMOs). There are two reasons why I bring up this point. First, as Jonathan Haidt and others have pointed out, Western higher education, especially in the Humanities and Social Sciences, has a strong leftwing bias in relation to the political orientation of the general population. Philosophy in general and social epistemology, in particular, are no exceptions. As academics working in these fields, we need to be careful not to reproduce and perpetuate this bias in our own work (I say this to myself as well, of course). Second, and this may be even more relevant from the authors' particular perspective, I doubt that it is even in their business interest to present BetterBelief in a way that may give the impression that it is a leftwing product (to dramatise the point), thereby excluding a significant portion of its potential market. For as I think we can all agree, there are reasonable and sincere people across the political spectrum. They might not share the same worldview, and their value systems may differ in striking ways, by nature or nurture, and yet they may very well be united in a common quest for better beliefs.

16d S. Kate Devitt, Kerrie Mengersen, Tamara R. Pearce and Alok Kumar Chowdhury's Response to Commentaries

de Ridder is concerned that BetterBeliefs will attract biased and overconfident experts and users, becoming an echo-chamber, reinforcing biases and providing even more evidence for pre-existing beliefs. Similarly, Olsson worries that BetterBeliefs is 'a contributor and content-neutral platform for people and organizations for whom truth matters', without sufficient mechanisms for managing free speech, curation and moderation. These are real concerns—particularly for a public release of the platform. Notwithstanding this, the team has worked with quite a few datasets and users, and we suggest that there are a number of ways in which these potential biases are ameliorated in our platform. First, any use of BetterBeliefs by a group should be scrutinised for who is invited to participate and how they are both nurtured and regulated in the use of the platform. Just as AI systems are only as valuable as the datasets on which they are trained, the outputs of BetterBeliefs are only as good as the input. Second, unlike many similar platforms, the echo chamber effect can be moderated by mechanisms such as allowing each user one vote per hypothesis at a time t , restricting the display of cumulative scores for degree of belief and weight of evidence, and creating composites of similar hypotheses. Third, the requirement for evidence, and weighting of this evidence by independent and group means, provides a further barrier to bias. The team is also currently developing machine learning methods that can automatically identify and reduce biases. These include item-response models to adjust for differences in user responses and the difficulty of assessing pieces of evidence, and anomaly detection statistics identify extreme or unusual individual patterns of responses. These approaches, and the platform itself, remain an active work in progress and our team is open to feedback and suggestions about how to improve these approaches.

To date, every use of BetterBeliefs has been curated and users hand-picked. The first users were software engineers of a global travel company. When they encountered the requirement to add evidence for their suppositions, it was such an unusual professional request that some of them added a meme such as 'show me the evidence!' Even analytic employees are not used to being asked to justify their beliefs. On the flip

side, users have almost universally found the social-media-like functionality very intuitive, which is one less hurdle. Users of BetterBeliefs are selected to represent diverse types of individuals with wide-ranging views and analytic preferences. Limits in the inputs and methods for each deployment have been noted and reported for analysis and subsequent engagement. Rather than the platform being used on a scrolling, never-ending basis (like Twitter or Facebook); the team sees the platform being used more for professional events (e.g. workshops), for specific time durations (e.g. to problem-solve a specific challenge), for a particular group purpose (e.g. evaluating research ideas). In comparison to existing social epistemic tools, we believe that the Better Beliefs platform offers new functionalities, for example, from a political perspective BetterBeliefs provides more nuance and justification than public surveys and online petitions—where bias to agree with one's peers may overrule critical thinking.

Olsson raises the Academic challenge of BetterBeliefs withholding mathematical details of our Degree of Belief (DoB) and Weight of Evidence (WoE) algorithms (while noting the commercial imperative to do so). We do not wish to dwell on the issue, but we do wish to acknowledge it. Each member of the BetterBeliefs founding team is a researcher—from Distinguished Professor Kerrie Mengersen, Dr Kate Devitt, Dr Alok Chowdhury and PhD Candidate Tamara Pearce. The underpinning methods are commercial in confidence for the time being. The algorithms are the emergent product of a transdisciplinary reckoning of normative social epistemology (philosophy), user-centred design (design), organisational decision making (business innovation) and the mathematical realisation of these norms (statistics and computer science). We note that business norms contrast with the virtue of openness and review that authors in this volume (including ourselves) generally endorse. The vice of secrecy must work hard to produce societal goods. That is to say, if algorithmic secrecy is required to maintain and sustain the platform that produces high-value social goods, is it justified? Or is individual or group ownership and profit of a platform such as ours unethical? Should a platform, if it brings genuine gain, be owned and operated by social collectives or governments rather than businesses? While our team operates the platform, we believe in the adaptive and risk-managed regulation of emerging technologies (Mandel 2009; Roca et al. 2017), including our own.

Finally, Olsson states that 'the platform is perhaps not so much an academic achievement as it is an attempt to bring social epistemology to the world of organizations and enterprises, a most laudable goal as I see it'. We appreciate Olsson's endorsement, but we would like to push back a little on what is considered 'Academic', as it is directly relevant to the main theme of the book—namely collective intellectual endeavour. Our team is transdisciplinary and impact-minded; drawing on a wide

range of Academic literature in business, mathematics, design, philosophy, decision science and cognitive science in order to create something genuinely new and intellectually justified. It is truly Academic. However, there is no doubt we have faced an uphill battle in Academia! Non-traditional outputs, such as decision support tools are not valued like journal articles or books in Academic performance metrics.

Transdisciplinarity re-imagines research disciplines and the possibilities for combining them rejecting disciplinary ‘silos’ (Broto et al. 2009; Lawrence & Després 2004), seeking to solve real-world, complex problems, assembling new approaches from scratch, using materials from existing scholarly disciplines for new purposes as well as reducing the gap between the researched and the researcher. However, the potential of transdisciplinary research takes additional time and resources to conduct the work, risk of miscommunication, differing priorities and lagging KPIs within research institutions to foster cross-silo research. Truly transdisciplinary research is frequently penalised in Academic literature, accused of not pushing foundational matters in one discipline or another by reviewers. There is a growing research methodology literature relevant to social epistemology that examines transdisciplinary methods, both their advantages for Universities in producing real-world solutions to highly complex global concerns and wicked problems (Bernstein 2015); and their pitfalls in dragging down the Academic careers of the individuals who embark on them. It is perhaps unsurprising that three of the four co-founders of BetterBeliefs have progressed careers in paid positions in research-friendly organisations outside of Academia (while maintaining close ties with research institutions). We believe that the future of social epistemology needs to continually engage experts, end users and diverse stakeholders both within and outside of the Academy.

References

- Bernstein, J. H. (2015). Transdisciplinarity: A review of its origins, development, and current issues. *Journal of Research Practise*, 11(1). Article R1.
- Broto, V. C., Gislason, M., & Ehlers, M. H. (2009). Practising interdisciplinarity in the interplay between disciplines: Experiences of established researchers. *Environmental Science & Policy*, 12(7), 922–933.
- Lawrence, R. J., & Després, C. (2004). Futures of transdisciplinarity. *Futures*, 4(36), 397–405.
- Mandel, G. N. (2009). Regulating emerging technologies. *Law, Innovation and Technology*, 1(1), 75–92.
- Roca, J. B., Vaishnav, P., Morgan, M. G., Mendonça, J., & Fuchs, E. (2017). When risks cannot be seen: Regulating uncertainty in emerging technologies. *Research Policy*, 46(7), 1215–1233.

17 Measuring Social Epistemic Virtues

A Field Guide

Marco Meyer

1 Introduction

When pursuing my PhD in economics at the University of Groningen, I was lucky to be handed the opportunity to field survey questions to a representative panel of Dutch households. The responses would be linked to a wealth of data about living conditions, employment and financial data, as well as health and psychological data. I worked on financial literacy and financial advice at the time. I was also fascinated by the idea of epistemic virtue. In particular, I wanted to find out whether epistemic virtues really were as important for the ability of people to gain knowledge and understanding as some applied epistemologists claimed—and others contested. Combining these two areas yielded an obvious research question: Are epistemically virtuous people more financially literate?

Surveys for measuring financial literacy were established and readily available. By contrast, I had no idea how to measure epistemic virtue. I was naïve enough to assume that I could solve this problem in an afternoon. Surely there would be a validated psychological survey instrument that measured epistemic virtue! My high expectations, though, were crushed. I still hoped I could quickly cobble together questions from existing instruments when I picked up a book about scale development (DeVellis 2016). In the introduction, Robert DeVellis warns social science researchers—like me—against simply using items that “look right” and relying on “existing instruments of questionable suitability.” As an alternative, he proposes that researchers should validate their own scale using psychometrical methods.

This afternoon marked the beginning of what became a sustained attempt to measure epistemic virtue (and vice) to answer questions in applied epistemology. After many false starts and failed attempts, and with Boudewijn de Bruin and Mark Alfano as collaborators, I eventually ended up validating a scale to measure epistemic vice (Meyer et al. 2021b). In the meantime, the field developed considerably, leaving researchers today with several options to measure epistemic virtue. This creates the challenge of selecting an appropriate instrument for the task. It also raises the question of what questions in applied epistemology can be solved using survey-based methods.

This chapter provides a “field guide” to measuring epistemic virtue using survey instruments, emphasizing *social* epistemic virtues. In an age of social media full of misinformation, social epistemic virtues are particularly important. In the Netflix documentary *The Social Dilemma* on the role of social media platforms in democracy, the former Google Design Ethicist Tristan Harris captures the challenge social media platforms pose very pointedly: “If we cannot agree on what is true, we are toast” (Orlowski 2020). Alongside an impressive cast of activists, academics, and wizards from the tech industry, Harris worries that their ads-based business model has nudged social media platforms to build their products to grab as much of our attention as possible. While we might use social media to learn about the world and form our views, the business incentive for social media platforms is to keep us hooked, not to educate us about what’s true. Unchecked by social epistemic virtue, we will connect to people who believe what we believe and build an echo chamber around us. As a result, we lap up content that confirms what we already believe.

Measuring social epistemic virtues poses some of the same and some special challenges as measuring epistemic virtue in general. In discussing these challenges and how to approach them, I proceed as follows. In the first section, I discuss whether and how social epistemic virtues can be measured in the first place. The answer to this question is not obvious, hinging on your conception of social epistemic virtue. I also discuss how you can approach measuring social epistemic virtues. Along the way, I consider a number of challenges for measurement: the role-relativity of social epistemic virtue; the challenge of “stealthy” virtues; and challenges in defining and individuating social epistemic virtues. I argue that all of these challenges can be overcome with thoughtful measurement strategies. In the second section, I discuss to what extent survey measures of social epistemic virtue can help answer questions in applied epistemology. I argue that empirical studies play an important role in informing foundational questions about social epistemic virtue. However, good measurement instruments for social epistemic virtue are a critical missing ingredient for designing studies that directly test key assumptions in important debates in virtue epistemology. In particular, I discuss to what extent studies can help to address the “situationist challenge” that social epistemic virtues do not explain knowledge acquisition. I argue that, in addition to sound measurement, we also need sophisticated study designs to identify causal relationships between social epistemic virtue and the quality of your epistemic network, rather than mere correlations.

2 Is the construct of social epistemic virtues measurable?

Any attempt at measurement should start from a clear definition of the construct to be measured. I will here consider social epistemic virtues

as the subset of epistemic virtues aimed at monitoring, adjusting and ameliorating one's epistemic environment, both for one's own sake and for the sake of others.

Consider social epistemic virtues concerned with *monitoring* your epistemic environment. If three of your friends endorse the same view, is it because they have researched the topic independently or is it because they amplify the same piece of misinformation a common friend of theirs has shared? Reflecting on such issues is critical to avoid believing fake news and spreading rumor. If you are virtuous in monitoring your social epistemic environment, you will double check the trustworthiness of a piece of content before you share it, particularly if it confirms what you already believe.

Consider social epistemic virtues concerned with *adjusting* your epistemic environment. These virtues concern your ability to put information about your epistemic environment to good use when forming beliefs. For instance, talking to more of your British friends about how they will vote in the Brexit referendum might lead you astray if most of them are from university towns. Social epistemic virtue manifests itself in weighing signals according to the partiality of your epistemic network and forming your beliefs accordingly. You can also display virtue in adjusting your epistemic environment by resisting to rely primarily on a news stream algorithmically curated by your friends. Max Hawkins, a former Google engineer, built an app to randomly select public Facebook events nearby (Spiegel & Rodriguez 2017). To break out of his bubble, he would attend the event that the app randomly selected. Most of the events Max would not have chosen to attend, from acro yoga to a community center pancake breakfast.

Finally, consider the ability to *improve* your epistemic environment. For instance, you can display this virtue by making a conscious effort to understand the other side of an issue by connecting with people and news sources you disagree with. A virtuous person would take steps to ameliorate their social network by including trustworthy sources on issues they care about. For instance, in the current global pandemic, a virtuous person might start following epidemiologists on Twitter. You can also help others in your network see the other side of an issue, for instance by suggesting articles exposing them to different views.

3 What kind of entities are social epistemic virtues?

With an intuitive understanding of social epistemic virtue under our belt, let's turn to the questions of whether and how it might be measured. To measure social epistemic virtue, we need to know where to look. What kind of thing is an epistemic virtue anyway, and where does it reside? Virtue epistemologists disagree in numerous philosophically interesting

ways about these questions. Which position you take has profound implications for whether epistemic virtue can be measured, and if so, how.

For a start, there are two “camps” that locate epistemic virtue very differently. Reliabilists maintain that epistemic virtues are faculties such as perception, induction, and memory (Sosa 2000). By contrast, responsibilists maintain that epistemic virtues are motivational dispositions to act in characteristic ways (Baehr 2011; Montmarquet 1993; Roberts & Wood, 2007; Zagzebski 1996). Different concepts of that which is to be measured require different approaches to measurement. Whereas reliabilists will investigate the quality of your eyesight or memory, responsibilists will attend to your character traits.

Another fundamental fault line is between realist vs. antirealist conceptions of epistemic virtue. Realists maintain that epistemic virtues are properties of individuals, not unlike height. For the realist, epistemic virtues are dispositions to act grounded in a person’s character. On this view, epistemic virtue can be measured by observing individuals. Because epistemic virtue only shows indirectly through the beliefs people hold and their actions, it is trickier to measure than visible properties like height. Yet introspection and observation of relevant actions are reasonable starting points for observing epistemic virtue on the realist’s view.

By contrast, anti-realists believe that epistemic virtues are attributions with self-fulfilling properties (Alfano 2013). For instance, by calling someone open-minded, you might influence them to act in an open-minded manner. Hence on the antirealist view, epistemic virtues reside in shifting social attributions, rather than in properties of individuals. On this account, measurement of epistemic virtue would be most directly approached by evaluating attributions of epistemic virtue, rather than a person’s character.

The survey instruments I consider here resonate best with a realist responsibilist account. Proponents of this view locate epistemic virtue in dispositions pertaining to a person’s character. These dispositions produce actions characteristic of the epistemic virtue in question, for example, sustained questioning in the case of inquisitiveness. However, antirealists might also embrace measurement using self-report surveys. Self-report surveys are well-suited to capture the readiness of people to attribute epistemic virtue to themselves. From an anti-realist perspective, such self-attributions of epistemic virtue can drive behavior.

You may wonder how epistemic virtues practically propel action. There is a sophisticated debate about whether epistemic virtues require good motives or not—with Zagzebski (1996), Tanesini (2018) and Battaly (2015, 2017) in favor, and Cassam (2016, 2019) against. Beyond discussion of the role of motivation, philosophers rarely say much about the psychological mechanisms that underlie epistemic virtue. An exception is Nancy Snow, Jennifer Cole Wright and Michael Warren (Snow

et al. 2020; see also: Jayawickreme & Fleenor 2017). They propose to think about virtues in general, moral or epistemic, in terms of the psychological theory of whole traits.

This framework is called *whole* trait theory because it unites a descriptive and an explanatory account of traits. Descriptively, whole trait theory identifies the degree to which you possess a trait by the frequency with which you display trait-appropriate responses. For instance, someone who is inquisitive should consistently and habitually fire questions at people who know something she does not. Note that this gives us a great hook for devising measurement approaches. How does whole trait theory explain the ability of virtues to trigger behavior? The theory breaks down the capacities someone needs to possess to act in trait-appropriate ways.

First, they need the capacity to perceive trait-relevant stimuli as relevant to the trait. For instance, an inquisitive person would note the mention of a theory they have not yet encountered as a prompt for asking questions.

Second, people need intermediate systems composed of belief and knowledge structures, motivational states such as desires, attitudes and cognitive schemas, as well as motivational states. For instance, someone who is inquisitive would experience a desire to ask questions when encountering a new theory, and they would be able to devise a battery of questions to test the theory and integrate it with what they already know.

This way of thinking about epistemic virtues is well suited to making them measurable because it pairs a psychologically plausible explanation of how epistemic virtues are psychologically “realized” with a descriptive account that makes measurement straightforward. Measures should seek to estimate in what proportion of virtue-relevant situations people behave virtuously.

4 Social epistemic virtues vs. epistemic outcomes

Why not just measure epistemic outcomes? For instance, we could assess the quality of someone’s epistemic environment by analyzing how homogeneous or otherwise their social network is. Similarly, we might survey them to assess whether they are prone to conspiracist beliefs or buy into Covid-19 misinformation. It is important to realize that in many use cases, this is not a good way to measure social epistemic virtue, because it conflates the epistemic outcomes we want to explain with the causes of these epistemic outcomes. Everyone agrees that people differ in knowledge and understanding. What is usually at issue is whether it is social epistemic virtue that explains these differences, in contrast to circumstance. Inferring epistemic virtue from epistemic outcomes begs the question of what really caused these outcomes: features about me,

or rather features about the situation I am in. My epistemic environment might be more heterogeneous because of my epistemic virtue, or simply because I live in a more heterogeneous area. Therefore, studies of social epistemic virtue typically need two components: First, a measure of virtuous behavior and attitude to capture the extent to which respondents display social epistemic virtue. Second, a measure of epistemic outcomes to assess whether social epistemic virtue is associated with better epistemic outcomes.

5 The challenge of role-relativity of social epistemic virtues

Whole trait theory has it that the measure of social epistemic virtue should estimate in what proportion of virtue-relevant situations people behave virtuously. This raises the question of how to identify virtuous responses. If we take our lead from the whole trait theory, a natural way to proceed would be to craft a test. The test would present respondents with a description of a situation, including virtue-relevant stimuli. Respondents could be asked how they would respond in these circumstances. Their response would be scored according to how close it comes to the virtuous action.

You won't be surprised to hear that nobody has attempted this way of measuring virtue, for two reasons: one more practical, the other more philosophical. The practical reason is that it is a bit of an embarrassment for any researcher to decide what epistemically virtuous action is. After all, virtue is an excellence, and it is supposed to be rare. Who am I, you might ask, to claim this excellence? The philosophical reason is that on the Aristotelian conception of virtue, it is a mean between extremes. For instance, inquisitiveness may be considered a mean between indifference and obsessive nosiness. Aristotle, for one, thought that the right mean must be determined relative to one's role in society. My choice of example already betrays my trade. Inquisitiveness is particularly relevant for researchers, whereas for, say, a TV host, epistemic justice may be a more salient virtue. Aristotle's point goes even further. He claims that the virtuous response for a researcher and a politician when encountering a new theory may well be different. Consider recent speculation that Covid-19 is a bioweapon manufactured in China. A virtuous researcher may manifest inquisitiveness by engaging with the theory, searching for confirming and disconfirming evidence. By contrast, a politician has to take into account the impact of raising the possibility publicly. Inquisitiveness might manifest for the politician in following up on the findings of researchers on the matter, rather than in engaging with the theory themselves. If you accept Aristoteles's version of the doctrine of the mean, there will not even be a single right response to a given situation, regardless of the role respondents have in society.

Survey instruments measuring epistemic virtue typically address these two problems by providing items that are quite abstract. For instance, the following is an item used by one scale measuring epistemic virtue to gauge the extent to which respondents are “engaged,” defined as being motivated to investigate things they do not yet understand (Alfano et al. 2017): “I enjoy reading about the ideas of different cultures.” In general, it seems appropriate to score people agreeing with this statement as motivated to investigate things they do not yet understand. Picking up books about foreign cultures is one of the things people might well do if they are so motivated. At the same time, anthropologists would likely hold themselves to a higher standard to agree to this statement than mathematicians would. That is a good thing, too, because it reflects the insight that the virtuous mean differs according to the role you have in society. Moreover, this approach sidesteps the problem of laying down what the virtuous response is by sticking to generalities.

6 The challenge of stealthy virtues

Note, though, that these advantages come at a price. We are no longer judiciously counting the proportion of trait-appropriate responses to trait-relevant stimuli, as whole trait theory would have us. Rather, we leave respondents discretion in making judgments about the extent to which they display a virtue. This is particularly problematic with regard to such epistemic virtues that are “stealthy.”

A trait is stealthy if possessing the trait stands in the way of knowing that you have the trait (Cassam 2015). Cassam suggests the vices of closed-mindedness and prejudice as candidates for stealthy traits. Consider the vice of closed-mindedness. If someone is closed-minded in general, their mind may be closed to the idea that they are closed-minded. It may well take a certain degree of open-mindedness to diagnose oneself as closed-minded. Certain virtues might be stealthy, too. One feature of the truly humble may be that they do not think about themselves as humble (Driver 1989). The boastful, on the other hand, are unlikely to fully appreciate their lack of intellectual humility (Alfano & Robinson 2014). In effect, the pretentious as well as the self-depreciatory may well lack the self-knowledge necessary to answer questions on intellectual humility correctly.

The potential stealthiness of social epistemic virtues has two implications for their measurement. First, to the extent that social epistemic virtues are stealthy, it is difficult to detect them using self-assessment measures. Yet the detection of stealthy virtues is merely difficult, rather than impossible, even in the case of fully stealthy virtues. The reason is that measures relying on self-assessment typically do not require respondents to attribute virtues or vices to themselves. Rather, researchers will ask about characteristic behaviors and attitudes of people possessing a

virtue. Someone who is closed-minded but would not describe themselves as such might still agree that they show behavior that is characteristic of the closed-minded, such as preferring a stable world view to constantly scrutinizing core beliefs. One test for whether a particular vice is stealthy is to compare respondent's self-assessment with an assessment of them by others. Alfano et al. conducted a study comparing self-ratings with ratings by informants on an intellectual humility scale and found positive correlations (Alfano et al. 2017, 12ff.). This result suggests that their humility scale picks up on traits that are not completely stealthy. Another test is to analyze correlations between self-assessments of epistemic virtue and epistemic outcomes. Bracketing the possibility of an underlying common cause for self-ratings and epistemic outcomes, such positive correlations would suggest that the epistemic virtues under consideration are not entirely stealthy.

A second implication for measurement is that it is difficult to infer the relative importance of epistemic virtues from self-assessment instruments. This is because epistemic virtues may be stealthy to different degrees. Suppose self-assessment measures of social epistemic virtue consistently find that, of two virtues, measures of one are more strongly associated with epistemic outcomes than measures of the other. If we take the possibility of stealthy virtues seriously, we cannot infer that one virtue is more strongly associated with epistemic outcomes than the other, because an alternative explanation is that the one virtue is less stealthy than the other.

7 The challenge to define social epistemic virtues

The ways philosophers often taxonomize social epistemic virtues do not fit easily with the requirements of scale construction. Consider the following list of epistemic virtues that Linda Zagzebski references (Zagzebski 1996): sensitivity to detail; open-mindedness in collecting and appraising evidence; fairness in evaluating the arguments of others; intellectual humility; intellectual perseverance; diligence, care and thoroughness; adaptability of intellect; being able to recognize reliable authority; insight into persons, problems, theories; the social virtues of being communicative; including intellectual candor and knowing your audience.

Zagzebski makes clear that she does not intend to give a comprehensive catalog of epistemic virtues. Nor are the epistemic virtues considered here clearly delineated from one another. For instance, sensitivity to detail and insight into persons, problems and theories clearly overlap, as do adaptability of intellect and open-mindedness. Moreover, some of these epistemic virtues appear to be moral virtues, too. For instance, fairness in evaluating the arguments of others can be seen as an instance of the moral virtue of fairness.

Yet psychological scale development requires definitions that are mutually exclusive. To see why, we need to understand the relationship between social epistemic virtues and the survey items that seek to measure them. Just like many other psychological constructs, social epistemic virtues cannot be directly observed. Rather, they are what psychologists call a latent trait. Latent traits are unobservable characteristics that cause observable behavior. For instance, the personality trait of extroversion cannot be observed directly. However, someone who is extroverted would particularly enjoy spending time in groups, whereas an introvert would enjoy spending time in crowds less. Survey items attempt to elicit endorsement from respondents showing these behaviors. For instance, extroverts should be more likely to endorse a statement like “I enjoy myself a lot at a lively party.”

For this methodology to work, everything hinges on items reflecting the latent traits they are meant to measure. Finding good candidate items is therefore the foundation of scale development. Researchers also need a rigorous and transparent process to show that items reflect the constructs they seek to measure. This is usually achieved by inviting a group of subject-matter experts to rate items according to how well they reflect the respective constructs. To make these decisions, subject-matter experts need sharp definitions to judge items against.

The reflective relationship between items and constructs also explains why definitions should be mutually exclusive. Suppose you want to use a scale to determine which social epistemic virtues explain the ability to detect fake news. To investigate this question, you need a scale that allows you to distinguish between distinct social epistemic virtues. If a given item reflects several social epistemic virtues, it will be impossible for you to distinguish which virtues respondents endorsing the item display. Consider the survey item “I like to complete my tax return well before the deadline.” This is a bad item for a personality survey, because endorsement of the item may reflect both conscientiousness and neuroticism.

8 The challenge to individuate social epistemic virtues

How do we determine whether two epistemic virtues are different in the first place? The philosophical literature makes several proposals. Zagzebski argues that epistemic virtues should be individuated by their proximate motivations (1996). Baehr has proposed to individuate epistemic virtues by the challenges to inquiry they are designed to overcome (2011). Scale construction approaches the individuation of traits from yet another perspective. The criterion for individuating traits is psychological. It may be possible to draw a conceptual distinction between two candidate virtues. If, however, these two candidate virtues always occur together in people, they would not be psychologically distinct.

There is no guarantee that virtues that are conceptually distinct are also psychologically distinct. In fact, the doctrine of the unity of the virtues provides some reason to doubt that they are. The unity of the virtues is the doctrine that the virtues are interdependent, such that a person cannot have any of the virtues without having all others (Wolf 2007). The doctrine goes back to Aristotle (Aristotle 2009, 1145a1–1145a2). According to the doctrine, we might be able to construct conceptually distinct intellectual virtues. But it would be either psychologically or conceptually impossible to display some of these virtues without displaying the others. In other words, the intellectual virtues may be conceptually distinct but would not be psychologically distinct.

If epistemic virtues would really form such a strong unity, this would lead to an inability of standard techniques of scale development such as factor analysis to distinguish epistemic virtues at all. Factor analysis (and its close cousin structural equation modeling) are statistical techniques used to establish the dimensionality of a construct and to select items (DeBode et al. 2013; DeVellis 2016). Factor analysis is a statistical method that extracts common factors explaining covariation between items. Factor analysis looks for regularities in linear combinations of scale items. For each set of items that are commonly scored similarly by participants, it extracts a factor. Ideally, factor analysis yields a factor for each of the social epistemic virtues your scale is attempting to measure. Each item can then be described by how strongly it varies with each of the extracted factors. Items that tap into the same social epistemic virtues should strongly load on the same factor and not display large loadings on other factors.

Consider measures of personality. The “Big-Five” inventory of personality traits was arrived at by testing many respondents on thousands of items spanning the whole domain of personality descriptions. Factor analysis reveals that there are underlying patterns behind these many items, at least among contemporary Westerners (Henrich et al. 2010), such that people score similarly on certain clusters of items. These clusters received the famous labels agreeableness, extroversion, open-mindedness, emotional stability and conscientiousness.

Note that according to this methodology, what counts as a distinct personality trait is determined by analyzing which traits people exhibit independently from each other. If, contrary to fact, every agreeable person was also conscientious and vice versa, factor analysis would subsume agreeableness and conscientiousness under the same category. But this is precisely what the doctrine of the unity of the virtues maintains with regard to virtues: possession of the virtues co-varies regardless of conceptual distinctions between them. If correct, factor analysis would not be able to pick up these differences (Peterson 2017). Hence if the doctrine of the unity of the virtues were correct, it would not be possible to distinguish epistemic virtues using this methodology.

Moreover, since according to the doctrine the possession of virtues covaries, items pertaining to one virtue would show large cross-loadings on other virtues. Large cross-loadings count against the inclusion of these items in the final scale according to standard methodology, because they are perceived to indicate that the item does not measure just one but several constructs. But according to the doctrine of the unity of the virtues, it would appear that we should expect to find cross-loadings even with items that tap into one specific virtue. Hence this methodology might lead us to abandon items that capture an epistemic virtue well because the item also loads on other factors.

9 How to measure social epistemic virtues

The list of challenges discussed above may look daunting. But all of these challenges can be overcome by constructing instruments carefully, while staying aware of their limitations in interpreting survey data. The role relativity of social epistemic virtue can be addressed in two ways. First, by crafting survey items that are general enough for people to respond against the backdrop of their role, while being specific enough to yield information about the trait in question. A general strategy for crafting such items is to stay as close to the definition of the item as possible. Second, by developing role-specific measures: tailoring measures to roles makes them less widely applicable, but it allows researchers to ask targeted questions that may yield more accurate measurements in the relevant populations.

The importance of stealthiness is best established in practice. As we will see below, epistemic virtues do not appear to be so stealthy that they cannot be measured using self-report measures at all. The likely effect of some stealthiness is that self-report surveys will underestimate the impact of social epistemic virtues on epistemic outcomes. Moreover, the possibility that different virtues are stealthy to different degrees raises a flag for the interpretation of survey results. We should not infer the relative importance of social epistemic virtues from the amount of variance they respectively explain in epistemic outcomes. The doctrine of the unity of the virtues is not a showstopper, either. If epistemic virtues really are unified as the doctrine suggests, this would merely imply that we cannot differentiate epistemic virtues. That still leaves open the possibility of measuring the impact of epistemic virtues as a whole. But we should not merely accept the doctrine. It is plausible that while social epistemic virtues covary with each other and other epistemic virtues, there is also some degree of independence. To what extent virtues can be differentiated empirically is best determined in the process of crafting measurement instruments. This work needs to start from clear and mutually exclusive conceptual distinctions between social epistemic virtues.

What do you practically do if you want to measure social epistemic virtue? One option is to find a psychological scale that measures a

construct that is “close enough” to the precise construct you want to measure for the purposes of your study. For instance, it might be that measures of open-mindedness capture important aspects of the social virtues associated with monitoring one’s epistemic network.

To my knowledge, there is no measure dedicated specifically to measuring social epistemic virtue. If you want to measure social epistemic virtue using a survey directly, this currently only leaves the option of validating your own instrument.

I recommend a three-pronged strategy to develop your own survey: work through a book on scale development, such as DeVellis (2016); study validation papers for scales in a similar area to inform your own design such as Alfano et al. (2017) and Krumrei-Mancuso and Rouse (2016); and find a psychometrician to advise you along the way.

Perhaps the most consequential part of scale development comes right at the beginning, when establishing the definitions of the constructs that you wish to measure. It is at this point that you decide what you consider the domain of social epistemic vice to be, for instance. Kidd has argued that what we considered epistemically virtuous and vicious has changed dramatically over time (2018). Curiosity, for instance, has evolved from a trait considered an epistemic vice in the Middle Ages to an epistemic virtue from the age of enlightenment on.

Once you have settled on clear definitions of the constructs you want to measure, validating the scale involves a series of validation steps.

I will illustrate the steps using the example of the epistemic vice scale that I validated with Boudewijn de Bruin and Mark Alfano.

Drafting. We initially drafted more than 300 items, none of which made it into the final version of the scale in its original form. We developed items based on the definitions of our constructs, as well as examples from the literature on epistemic virtue. We also reviewed existing related scales to take inspiration for our items.

Expert Validation. We invited experts working on intellectual virtue and vice to rate items according to how well they reflect our definitions of epistemic vices, which led to a fundamental revision of the item pool.

Discrimination analysis. We asked a convenience sample of 20 people to rate items depending on how well they reflect each of the intellectual virtues. This step helped us in identifying items that tap into exactly one of the virtues that we seek to measure and led to a further revision of the item pool.

Exploratory factor analysis. We recruited participants (on Amazon Mechanical Turk) to respond to all items in the revised item pool. We then conducted an exploratory factor analysis on the items. The analysis revealed that a subset of the items has a clean factor structure. We optimized scale length and ended up with a scale consisting of two subscales and ten items.

Confirmatory factor analysis. Using responses from a further study administering the items identified during the explanatory factor analysis, we conducted a confirmatory factor analysis to respond to the items identified during the exploratory factor analysis. We found that the factor structure remained stable in the confirmatory analysis.

Convergence and Divergence analysis. While gathering data for the confirmatory data analysis, we also asked participants to fill in a number of related scales and demographic information to establish whether the construct the scale taps into is distinct from other psychological measures, such as dogmatism or personality. We found that the scale indeed taps into a distinct construct.

Construct validity. We also conducted a study to test whether the scale was related to outcome measures such as the Covid-19 misinformation items and the fake news items. I will describe this study more fully below.

There are other desirable steps in validating a survey instrument, such as testing the stability of responses over time (test-retest validity) and testing whether self-reports coincide with reports from third parties. Yet following these steps goes a long way to ensure that your survey instrument is of high quality. While it is a significant undertaking, the benefits are substantial. Once you have a validated instrument at your disposal, you have the central ingredient for designing studies that can help inform the most contentious issues in virtue epistemology.

10 Can measurement help to solve questions in applied epistemology?

Assuming that you got hold of a measure of social epistemic virtue, what to do with it? In this section, I first discuss the role of empirical studies for settling debates in applied epistemology. I illustrate the importance of instruments to measure epistemic virtue by discussing their role in informing the situationist debate. It will turn out that while good measurement is necessary, it is not sufficient for making progress on the situationist debate. We also need sophisticated research designs to harness the power of good measurement. Then I apply the learnings to how measures of *social* epistemic virtue can be used to inform debates in virtue epistemology.

There is broad agreement between virtue epistemologists that empirical studies are needed to resolve important philosophical issues in the field (Alfano 2013, 118; Cassam 2016, 170; Zagzebski 1996, 309). Accordingly, researchers in the field routinely appeal to empirical work, primarily from psychology. Yet even the most relevant studies were often not designed to address the philosophical questions at stake. That contributes to disagreement over how to interpret the results of these studies for the purposes of philosophical issues. Reliable

instruments to measure social epistemic virtue are the first step to design studies that explicitly address key philosophical questions in virtue epistemology.

11 The role of empirical studies in evaluating the situationist challenge

Let's consider whether better measurement might help in making progress on the situationist challenge. Situationists about epistemic virtues challenge that epistemic virtues explain differences in knowledge acquisition. This is a fundamental challenge to virtue epistemology, because many virtue epistemologists agree with the following two claims. First, that knowledge is true belief, acquired through epistemic virtues. Second, most people know a lot of things. But these two thoughts are incompatible with the situationist's contention that most people's epistemic traits are not virtuous.

To make the claim that most people do not possess epistemic virtue, situationists typically appeal to research showing that people's epistemic actions are highly sensitive to situational influences. For instance, watching a funny video or eating a piece of chocolate has a surprisingly large influence on how creative people are (Isen et al. 1987).

The situationist challenge can be expanded to social epistemic virtues, for instance by showing that for most people, a small change in their epistemic network has a large effect on their beliefs. For instance, an experiment by Facebook showed that people receiving a call to vote on election day identifying concrete people from the recipient's friend network who had already voted were 2% more likely to affirm that they had voted than people who received the same message without their friend's pictures attached to it (Bond et al. 2012).

These studies attempt to infer the lack of explanatory power of epistemic virtue from the observed impact of a circumstantial factor on knowledge acquisition. This indirect strategy opens up several ways for proponents of epistemic virtue. One is to marginalize the importance of the virtue under consideration for knowledge acquisition. Another is to acknowledge the impact of situational factors on changes of epistemic conduct yet insist on the explanatory power of epistemic virtues for conduct that is sufficient for knowledge acquisition. Perhaps chocolate makes people even more creative, but this does not provide evidence for a lack of creativity absent a chocolate boost.

Absent a measure of epistemic virtue, there is no alternative to this indirect strategy. That can lead to an unsatisfying standoff between situationists and proponents of epistemic virtue. The proponent of epistemic virtue does not need to deny that situational factors have an influence on knowledge acquisition. All they need to maintain is that virtue is an important factor for explaining knowledge acquisition. Studies using the

indirect approach do not directly challenge this latter claim. However, mounting evidence about the importance of circumstance can nibble away at the credibility of epistemic virtue. At the same time, the indirect strategy holds out no hope for evidence in favor of the explanatory power of epistemic virtue. At best, such studies find no impact of a situational factor on knowledge acquisition, which does not amount to positive evidence in favor of epistemic virtue. As a result, proponents of virtue epistemology find themselves in a reactive position, forced to explain away ever new demonstrations of the effects of seemingly epistemically irrelevant situational factors. It is worth noting that this dialectic is not the result of the merits of either position. Rather, it is an artifact of an empirical strategy that does not use a direct measure of epistemic virtue.

12 How measures of epistemic virtue can help to inform the situationist debate

Given an instrument to measure social epistemic virtue, we can test the claim that differences in epistemic virtue explain differences in knowledge acquisition directly. Such studies can provide new evidence relevant to assessing the situationist challenge. If such studies show an explanatory relationship between social epistemic virtue and knowledge acquisition, this provides a positive reason to posit such virtues. If, however, studies fail to show an explanatory link, this provides reason to doubt the explanatory role of social epistemic virtues.

In its basic form, such studies need a metric for the relevant epistemic virtues, a metric for a measure of knowledge acquisition, and metrics for confounding factors. Let me illustrate this approach with a recent observational study of the relationship between epistemic vice in general and the readiness of people to endorse Covid-19 misinformation (Meyer et al. 2021a). I will consider implications of the design of studies that focus on *social* epistemic duties thereafter.

We administered a survey to 998 respondents from the United States. The survey had three parts. The first part is an instrument measuring epistemic vice that we had validated independently. The instrument consists of ten items that are answered on an agree-disagree scale. The responses of each participant are averaged across the ten questions, yielding an epistemic vice score.

The second part contains outcome measures measuring knowledge acquisition. The Covid-19 misinformation measure tests whether participants believe claims about Covid-19 that the WHO has identified as demonstrably false. A second measure tests whether participants find fake news articles about Covid-19 credible. The responses are aggregated into a Covid-19 misinformation score and a fake-news score, respectively.

The third and final part of the survey contains measures of potentially confounding factors, including political affiliation, religiosity,

demographic variables and a battery of other psychological measures, including measures of personality and dogmatism.

We found strong evidence to the effect that epistemic vice is associated with susceptibility to Covid-19 misinformation and fake news. In fact, the correlation between epistemic vice score and these outcome metrics turns out to be stronger than with political identity, educational attainment, personality, dogmatism—and any other competing explanation that we tested. What is more, epistemic vice explains a sizable chunk of additional variance in misinformation and fake news scores when controlling for all confounding factors in a regression model.

This type of evidence provides relevant data to the situationist debate. It establishes a positive link between self-reports of epistemic virtue and measures of knowledge acquisition in a domain of knowledge of the most acute general interest. Because of the large size of the effect we find, one might take these findings to provide evidence for the explanatory power of epistemic virtues.

13 We also need sophisticated study designs to settle debates in virtue epistemology

Yet this type of study does not provide a knockdown argument against situationism. There are at least two ways for situationists to resist the conclusion that epistemic virtue explains knowledge acquisition. First, proponents of situationism can question the direction of causation. Ours is an observational study and can as such only establish a correlation between the measure of epistemic virtue and the measure of knowledge acquisition. For all the study shows, the causal link could run from knowledge acquisition to self-reports of epistemic virtue/vice. This would be the case if people wrongly attribute epistemic virtue to themselves on the ground that they experience themselves as having a lot of knowledge. Thus the situationist may try to explain away the relevance of the evidence by appealing to the fundamental attribution error, i.e., the tendency to attribute behavior to character traits rather than situational factors (Harman 1999).

The challenge to establish the direction of causation cannot be fully addressed given the study design we used. But note that it is a perfectly general challenge to observational studies. There are numerous empirical research methods that help to determine the direction of causation, including randomized experiments, longitudinal studies, and regression designs using instrumental variables. All of these methods require strong instruments for measuring epistemic virtue. But these measures need to be embedded in sophisticated research designs.

Another way for situationists to resist the conclusion that epistemic virtues explain knowledge acquisition is to appeal to confounding factors. Confounding factors would be features that influence both responses on the epistemic virtue measure and responses to the respective outcome

measure. For instance, educational attainment might lead respondents to have more favorable views of themselves concerning their epistemic virtue and vice, as well as improve knowledge—without epistemic virtue and knowledge acquisition being causally related.

To address this possibility, studies should include measures of confounding factors and use a regression design to determine to what extent epistemic virtue explains variance in knowledge acquisition “over and above” potential confounding factors. Careful development of the measurement instrument pays off at this stage because it guides researchers towards an instrument that measures the construct of interest *and that construct alone*. By contrast, merely jotting down some items will likely result in an instrument that taps into some aspects of epistemic virtue/vice, but also plenty of related constructs besides. The price to pay for a construct that is not carefully validated is the risk of a high correlation not between the epistemic virtue and the constructs you treat as controls, but between your measure of epistemic virtue and control variables. As a result, epistemic virtue might appear to explain less additional variance in individual differences in knowledge acquisition than it otherwise would.

Hence good metrics combined with sophisticated research designs can help make progress in the situationist debate, which is probably the most fundamental controversy in virtue epistemology. However, the benefits of good metrics are not limited to the situationist debate. Other debates requiring research designs with good measurement instruments include whether (social) epistemic virtues are character traits or mere attributions; whether we have self-knowledge about (social) epistemic virtues and vices; how (social) epistemic virtues are acquired; whether different upbringings, teaching and training foster different (social) epistemic virtues; and which (social epistemic) virtues are most beneficial in particular contexts.

14 Implications for studying social epistemic virtues

One take-away from the previous discussion for the study of social epistemic virtues is that research designs that can inform debates in virtue epistemology typically measure three things: a measure of social epistemic virtue, an outcome measure and measures of control variables. I have emphasized the importance of using validated metrics to measure social epistemic virtue. A general measurement instrument for social epistemic virtue is still outstanding. Researchers need to either validate their own or make do with surrogates close enough to the construct they seek to measure for their research question.

However, researchers are in a good place when it comes to identifying outcome measures for studying social epistemic virtue. Online platforms like Twitter and Facebook provide an excellent window into our epistemic networks. While epistemic networks are difficult to chart offline, online platforms provide a wealth of detailed information.

Sullivan et al. have developed the groundwork for metrics to quantify the epistemic position of people in a network (Sullivan et al. 2020a, 2020b). They develop a formal definition of a person's epistemic position based upon three factors: the number of different viewpoints among a person's sources; the number of informants; and the degree to which informants are independent from one another. Based on this formalization, they propose a metric that aggregates these three factors into a single score. They also demonstrate that their metric can usefully be applied to data from Twitter.

This metric is an excellent candidate for outcome measures when studying social epistemic virtue. It is plausible that people who are good at monitoring, adjusting, and ameliorating their epistemic network would find themselves in a better epistemic position measured this way than people who lack at least one dimension of social epistemic virtue, other things being equal. Since the quality of one's epistemic network can be analyzed separately for different issues, the measure is also useful in studying to what extent social epistemic virtues are global or local, i.e., whether they apply to a narrow or a broad domain of issues.

One limitation of the metric is that it only captures the part of the epistemic network that is present on the given social media platform. As a result, studies will tend to underestimate the number of sources people have access to.

15 Conclusion

Social epistemic studies are still a new area of investigation in philosophy. That is reflected in the lack of dedicated measurement instruments. Early studies of online epistemic networks suggest that there are large individual differences in the quality of epistemic networks, concerning the number of sources, the independence of sources, as well as the diversity of opinions represented. Developing and validating proper measurement instruments is the first step for investigating whether and to what extent social epistemic virtues drive those differences. In addition to good measuring instruments, we also need sophisticated study designs to establish a causal—rather than merely correlation—relationship between social epistemic virtue and the quality of epistemic networks.

References

- Alfano, M. (2013). *Character as Moral Fiction*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139208536>
- Alfano, M., Iurino, K., Stey, P., Robinson, B., Christen, M., Yu, F., & Lapsley, D. (2017). Development and Validation of a Multi-Dimensional Measure of Intellectual Humility. *PLoS ONE*, 12(8), 1–28. <https://doi.org/10.1371/journal.pone.0182950>

- Alfano, M., & Robinson, B. (2014). Bragging. *Thought: A Journal of Philosophy*, 3(4), 263–272. <https://doi.org/10.1002/tht3.141>
- Aristotle. (2009). *The Nicomachean Ethics* (L. Brown, Ed.; D. Ross, Trans.; Revised edition). Oxford University Press.
- Baehr, J. (2011). *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford University Press.
- Battaly, H. (2015). Epistemic Virtue and Vice: Reliabilism, Responsibilism, and Personalism. In C. Mi, M. Slote, & E. Sosa (Eds.), *Moral and Intellectual Virtues in Western and Chinese Philosophy* (pp. 109–130). Routledge.
- Battaly, H. (2017). Testimonial Injustice, Epistemic Vice, and Vice Epistemology. In I. J. Kidd, J. Medina, & G. Pohlhaus, Jr. (Eds.), *The Routledge Handbook of Epistemic Injustice* (pp. 223–231). Routledge.
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-Million-Person Experiment in Social Influence and Political Mobilization. *Nature*, 489(7415), 295–298. <https://doi.org/10.1038/nature11421>
- Cassam, Q. (2015). Stealthy Vices. *Social Epistemology Review and Reply Collective*. <https://social-epistemology.com/2015/10/16/stealthy-vices-quassim-cassam/>
- Cassam, Q. (2016). Vice Epistemology. *The Monist*, 99(2), 159–180. <https://doi.org/10.1093/monist/onv034>
- Cassam, Q. (2019). *Vices of the Mind: From the Intellectual to the Political*. Oxford University Press.
- DeBode, J. D., Armenakis, A. A., Feild, H. S., & Walker, A. G. (2013). Assessing Ethical Organizational Culture Refinement of a Scale. *The Journal of Applied Behavioral Science*, 49(4), 460–484. <https://doi.org/10.1177/0021886313500987>
- DeVellis, R. F. (2016). *Scale Development: Theory and Applications*. SAGE Publications.
- Driver, J. (1989). The Virtues of Ignorance. *The Journal of Philosophy*, 86(7), 373–384. <https://doi.org/10.2307/2027146>
- Harman, G. (1999). Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error. *Proceedings of the Aristotelian Society*, 99, 315–331. JSTOR.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most People Are Not WEIRD. *Nature*, 466(7302), 29–29. <https://doi.org/10.1038/466029a>
- Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive Affect Facilitates Creative Problem Solving. *Journal of Personality and Social Psychology*, 52, 1122–1131.
- Jayawickreme, E., & Fleeson, W. (2017). Does Whole Trait Theory Work for the Virtues? In *Moral Psychology: Virtue and Character*, Vol. 5 (pp. 75–103). Boston Review. <https://doi.org/10.2307/j.ctt1n2tvzm.9>
- Kidd, I. J. (2018). Deep Epistemic Vices. *Journal of Philosophical Research*, 43, 43–67. <https://doi.org/10.5840/jpr2018431>
- Krumrei-Mancuso, E. J., & Rouse, S. V. (2016). The Development and Validation of the Comprehensive Intellectual Humility Scale. *Journal of Personality Assessment*, 98(2), 209–221. <https://doi.org/10.1080/00223891.2015.1068174>
- Meyer, M., Alfano, M., & de Bruin, B. (2021a). Epistemic Vice Predicts Acceptance of Covid-19 Misinformation. *Episteme*. <https://doi.org/10.1017/epi.2021.18>

- Meyer, M., Alfano, M., & de Bruin, B. (2021b). The Development and Validation of the Epistemic Vice Scale. *Review of Philosophy and Psychology*. <https://doi.org/10.1007/s13164-021-00562-5>
- Montmarquet, J. A. (1993). *Epistemic Virtue and Doxastic Responsibility*. Rowman & Littlefield.
- Orlowski, J. (2020). *The Social Dilemma*. Netflix. <https://www.netflix.com/de/title/81254224>
- Peterson, C. (2017). Exploratory Factor Analysis and Theory Generation in Psychology. *Review of Philosophy and Psychology*, 8(3), 519–540. <https://doi.org/10.1007/s13164-016-0325-0>
- Roberts, R. C., & Wood, W. J. (2007). *Intellectual Virtues: An Essay in Regulative Epistemology*. Oxford: Oxford University Press.
- Snow, N. E., Wright, J. C., & Warren, M. T. (2020). Virtue Measurement: Theory and Applications. *Ethical Theory and Moral Practice*, 23(2), 277–293. <https://doi.org/10.1007/s10677-019-10050-6>
- Sosa, E. 2000. Three Forms of Virtue Epistemology. In Axtell, G. (Ed.), *Knowledge, Belief and Character: Readings in Virtue Epistemology* (pp. 33–40). New York: Rowman & Littlefield.
- Spiegel, A., & Rodriguez, M. (2017, June 8). Eager to Burst His Own Bubble, A Techie Made Apps to Randomize His Life. *NPR.org*. <https://www.npr.org/sections/alltechconsidered/2017/06/08/531796329/eager-to-burst-his-own-bubble-a-techie-made-apps-to-randomize-his-life>
- Sullivan, E., Sondag, M., Rutter, I., Meulemans, W., Cunningham, S., Speckmann, B., & Alfano, M. (2020a). Vulnerability in Social Epistemic Networks. *International Journal of Philosophical Studies*, 28(5), 731–753. <https://doi.org/10.1080/09672559.2020.1782562>
- Sullivan, E., Sondag, M., Rutter, I., Meulemans, W., Cunningham, S., Speckmann, B., & Alfano, M. (2020b). Can Real Social Epistemic Networks Deliver the Wisdom of Crowds? In T. Lombrozo, J. Knobe, & S. Nichols (Eds.), *Oxford Studies in Experimental Philosophy, Volume 1* (pp. 29–63). Oxford University Press.
- Tanesini, A. (2018). Epistemic Vice and Motivation. *Metaphilosophy*, 49(3), 350–367. <https://doi.org/10.1111/meta.12301>
- Wolf, S. (2007). Moral Psychology and the Unity of the Virtues. *Ratio*, 20(2), 145–167. <https://doi.org/10.1111/j.1467-9329.2007.00354.x>
- Zagzebski, L. T. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge University Press.

17b Commentary from T. Ryan Byerly

Marco Meyer's chapter provides a "field guide" for using empirical methods to measure and study social epistemic virtues for philosophical purposes. After briefly introducing the concept of social epistemic virtues, Meyer describes how methods used in psychology to measure and study character traits can be fruitfully applied to social epistemic virtues, and he then goes on to explain how studies can be designed to use measures of social epistemic virtues to help answer questions of interest to philosophers.

I will offer two brief critical comments on Meyer's chapter. The first comment concerns the concept of social epistemic virtues. Meyer defines these as "a subset of epistemic virtues aimed at monitoring, adjusting, and ameliorating one's epistemic environment, both for one's own sake and for the sake of others." Here "epistemic environment" is best understood as a *social* epistemic environment constituted by a network of fellow inquirers in which one is a participant. Meyer's interest is in traits that involve attending to epistemically significant features of such environments, adjusting one's position in such environments when doing so will be epistemically beneficial, and endeavoring to enhance the epistemic features of these environments.

This way of conceptualizing social epistemic virtues gives us a place to start for purposes of developing measures of these and studying them empirically. Yet, at the same time, it leaves quite a bit of conceptual work to be done. In particular, it leaves unanswered the following questions: Are there specific epistemic virtues that are social epistemic virtues, while other epistemic virtues are not social epistemic virtues? Are all so-called "traditional" (Daukas 2019) epistemic virtues social epistemic virtues to some extent—perhaps when possessed "in their fulness" (Baehr 2011)? For instance, are open-mindedness or intellectual thoroughness or attentiveness social epistemic virtues? Put a bit differently, the question is whether the concerns for one's social epistemic environment that are definitive of social epistemic virtues are concerns distinctive of some subset of epistemic virtues, or whether all or many traditional epistemic virtues, when possessed in their fulness, might imply concerns of this sort.

A related question concerns whether we should think of this account of what social epistemic virtues are as having any special claim to correctly capturing what social epistemic virtues are—or whether there might be other kinds of epistemic virtues that are also reasonably thought of as social epistemic virtues, but that do not satisfy this account. I took it that Meyer’s proposal was largely stipulative, rather than aiming to offer an analysis of some sort of pre-theoretical concept of social epistemic virtues. Still, it may be worth noting that there are other candidates for epistemic virtues that are markedly social in some way, but perhaps not in the specific way Meyer has in mind. For instance, some of my own recent work has focused on “virtues of intellectual dependability” that distinctively involve a motivation to promote others’ epistemic goods (Byerly 2021). These certainly seem to be markedly social in a way, but they aren’t necessarily social in the way Meyer focuses on. Or, at any rate, if they are social in that way too, they are so only “in their fulness.” Are they still “social epistemic virtues”?

My second comment is concerned with some of Meyer’s remarks about how to develop measures of epistemic virtues. Meyer highlights very helpfully what we might think of as a kind of problem-facing philosophers who wish to study the traits that interest them using empirical methods. The problem arises because philosophers are often interested in rather subtle differences between traits. This leads to the proliferation of accounts of distinct virtues. A recently edited volume contains chapters on twelve different epistemic virtues, for instance (Battaly 2019). Yet, as these philosophers will quickly acknowledge, the traits overlap with one another in all kinds of ways. When it comes to empirical measurement, this creates a problem, because the higher correlated these traits are with one another, the less unique explanatory work they will be able to do. In fact, if they are highly-enough correlated with one another, then the standard methods such as factor analysis that Meyer recommends will suggest that the traits are not empirically distinct—that we should treat conceptually distinct virtues as in fact reflecting the same latent psychological construct.

I think this flags a serious concern for virtue epistemologists. Yet, at the same time, I think Meyer’s remarks overstate the challenge and especially what should be done to address it. In explaining the challenge, for instance, Meyer writes that “If a given item reflects several social epistemic virtues, it will be impossible for you to distinguish which virtues respondents endorsing the item display.” And, in recommending a solution, Meyer writes that “psychological scale development requires definitions [of traits] that are mutually exclusive.” I think both remarks are too extreme.

First, it is possible for distinct scales to share one or more items in common, or to share items that are extremely similar, and yet for the scales themselves to measure empirically distinguishable constructs. I’ve

been dealing with just this kind of issue in some of my own research. In psychology, there are multiple, distinct constructs that all overlap in that they reflect a person's tendency to prioritize others' interests over their own interests, yet they differ from one another when it comes to *why* the person prioritizes others' interests (e.g., Helgeson & Fritz 1998; Wright et al. 2018). The similarity between these traits can be reflected by including some similar or even identical items in the scales. But the different motivations leading to these similarities will be reflected in other items, yielding scales that remain empirically distinguishable, despite having some overlap.

Second, the solution to the challenge does not require providing definitions of traits that are "mutually exclusive." To provide definitions of traits T1 and T2 that are mutually exclusive is to provide definitions such that a person cannot possess both T1 and T2. But this is unnecessary for scale development, and any personality psychologist would tell you so. Something weaker, instead, is true. It is fine to use a concept of an epistemic virtue that overlaps with concepts of other epistemic virtues for purposes of scale development. But, it is important to include in one's initial item writing a healthy dose of items that also reflect the distinctive features of the trait. We might think of this as "playing up" or emphasizing the distinctive features of the traits we're studying, while also acknowledging their similarities with other traits. There's always a risk that in using this procedure we develop measures of conceptually distinct virtues that are empirically indistinguishable. But, if we're going to be faithful to the philosophical conception of the traits we're studying, it's a risk we may be justified in taking.

References

- Baehr, Jason. 2011. *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. New York: Oxford University Press.
- Battaly, Heather, ed. 2019. *The Routledge Handbook of Virtue Epistemology*. New York: Routledge Press.
- Byerly, T. Ryan. 2021. *Intellectual Dependability: A Virtue Theory of the Epistemic and Educational Ideal*. New York: Routledge Press.
- Daukas, Nancy. 2019. "Feminist Virtue Epistemology." In *The Routledge Handbook of Virtue Epistemology*, ed. Heather Battaly, 379–391. New York: Routledge.
- Helgeson, V., & H. Fritz. 1998. A Theory of Unmitigated Communion. *Personality and Social Psychology Review*, 2(3), 173–183.
- Wright, J. C., T. Nadelhoffer, L. Thomson Ross, & W. Sinnott-Armstrong. 2018. Be It Ever So Humble: Proposing a Dual-Dimension Account and Measurement of Humility. *Self and Identity*, 17(1), 92–125.

17c Commentary from Alessandra Tanesini

Measuring Social Epistemic Virtues

In his contribution to this volume, Meyer provides a very helpful step-by-step guide on how to develop psychometric scales, and more specifically on how to produce these measuring instruments with respect to social epistemic virtues and vices. In this response, I focus on Meyer's definition of a social epistemic virtue, and on some of the obstacles with measuring these virtues so conceived. None of my worries constitute insurmountable problems for Meyer's approach but ultimately encourage a modification of the metaphysics of virtue he endorses in his chapter.

Meyer defines "social epistemic virtues as the subset of epistemic virtues aimed at monitoring, adjusting [to], and ameliorating one's epistemic environment, both for one's own sake and for the sake of others" (p. 3). There are several aspects of this definition that in my view require clarification. First, the notion of epistemic environment is on one plausible interpretation too capacious to play the role intended by Meyer since it would include the whole of the information available to the agent. Since Meyer's examples concern exclusively the reception of testimony, one might restrict epistemic environment to include only the judicious seeking and assessing of the testimony of agents offered in speech or print. So understood social epistemic virtues might be too narrowly circumscribed since the definition would seem to exclude other virtues of epistemic dependence. These could include the epistemic virtue of a good teacher that might aim to improve the epistemic character of the students, but also the virtues of a good testifier who is able to formulate her testimony in a vocabulary that is suitable to her audience. It would seem a stretch to think of these virtues as aimed at improving the epistemic environment.

Second, and relatedly, the idea that the monitoring, adjustment or amelioration is done for someone's sake needs some explication. I presume here that what Meyer's means is that the aim of these virtuous activities is the epistemic well-being or flourishing of agents. Such well-being should include knowledge acquisition, deeper understanding, and the

reduction of error. But, if my suggestion that the virtues of helping others to become better epistemic agents should be included among the social epistemic virtues is correct, the improvement of epistemic character is an additional end of virtuous activity.

Third, and finally, Meyer does not say whether aiming at the right goals and for the right reasons is sufficient to be virtuous or whether reliable success in ordinary circumstances is also required to be virtuous. Meyer rightly points out that measuring virtue is not the same as measuring epistemic outcomes (p. 6), since even reliable success could be due primarily to environmental factors or traits that are not virtuously motivated. In what follows I presume that virtue must reliably produce success; this presumption seems to fit best with Meyer's description of the kind of survey items used to measure virtuous behavior in self-reports.

These observations indicate that Meyer might have defined social epistemic virtues too narrowly. I suspect that the addition of virtues aimed at the improvement of epistemic character, and of the virtues of being receptive to the epistemic needs of other agents to the number of the social epistemic virtues might generate problems for Meyer's proposal to measure virtue using exclusively self-reports. The approach consists in presenting subjects with several statements describing preferences, opinions and behaviors to elicit whether in their view said statements accurately capture what they like, believe or do. Whilst this approach might be suitable to assess an agent's motivations and some aspects of their behavior, if we want to measure virtues that require a perceptive response to others' epistemic needs, it seems at least plausible that any good method of measurement must include third parties' reports. The need for such an addition is not an insurmountable obstacle, but it raises interesting questions about how to treat extensive inconsistencies between reports were these to occur.

Be that as it may, further problems arise if one adopts, as Meyer appears to do, the metaphysics of virtues as dispositions that is endorsed by Snow et al. (2019) when they propose to measure virtue by measuring the frequency with which an individual behaves virtuously in virtue-relevant situations (Meyer, 5–6). So understood virtues are dispositions that have eliciting conditions. When these obtain, the behavior characteristic of the virtue would be triggered. Virtue would thus be a psychological feature that is ontologically like solubility or fragility. These are characteristics of objects that have the propensity to dissolve or to shatter in specific sets of circumstances.

This dispositional "if virtue-eliciting situation then virtuous behavior" model is not suitable for several virtues. The point was originally made by Rees and Webber (2014) and bears repeating. There are virtues like integrity that should be manifest in all circumstances, rather than being elicited only by some triggering situations. There are also virtues

like generosity that partly consist in seeking, and creating, the situations relevant to their manifestations. These virtues have intellectual components and might be thought of as relevant to relations of interpersonal epistemic dependence.

In addition, the “if-then” dispositional account of virtue presupposes that virtues must have so-called high-fidelity. If, as Alfano (2013) has suggested, other virtues do not require that they are manifested in most cases of exposure to virtue-eliciting situation, then the if-then model of virtues as dispositions is ill-suited as a characterization of virtues of this low-fidelity kind. It is worth noting that several epistemic virtues have low fidelity. For instance, we are inclined to think of a person as open-minded if she is prepared, and able, to engage with varied opinions alternative to her own on at least some occasions. Hence, for instance, we do not think that the person who has some blind spots on a few circumscribed issues is closed-minded. The corresponding epistemic vice of closed-mindedness is instead, as one would expect, high fidelity (Cassam 2019, 45).

If applicable to several epistemic virtues, this consideration raises questions about the suitability of the “if-then” model of virtue to virtue epistemology. I have argued elsewhere that issues such as these support the development of a different account of virtue in terms of attitudes, rather than dispositions to respond to virtue eliciting situations (Tanesini 2021, ch. 3). That said, I believe that this objection can in practice be accommodated in the approach to measurement proposed by Meyer since as a matter of fact, it consists in the development of instruments to measure participants’ attitudes.

Consider, for example, the survey item “I enjoy reading about the ideas of different cultures” used by Alfano et al. (2017) in their instrument to measure intellectual virtue. This item is ill-suited to capture responses to triggering stimuli. Instead, it seems to measure attitudes to reading about novel ideas, which are manifested among other things in behavior that includes seeking to create opportunities to engage in the behavior that is characteristic of open-mindedness. If this is right, the problem I singled out here lies with a mismatch between philosophical theory and measurement methods requiring a change of theory rather than a change in approach to measurement.

References

- Alfano, M. (2013). *Character as Moral Fiction*. Cambridge: Cambridge University Press.
- Alfano, M., Iurino, K., Stey, P., Robinson, B., Christen, M., Yu, F., & Lapsley, D. (2017). Development and Validation of a Multi-Dimensional Measure of Intellectual Humility. *PLoS ONE*, 12(8), e0182950. doi:10.1371/journal.pone.0182950.

- Cassam, Q. (2019). *Vices of the Mind*. Oxford: Oxford University Press.
- Rees, C. F., & Webber, J. (2014). Constancy, Fidelity and Integrity. In S. Van Hoof (Ed.), *The Handbook of Virtue Ethics* (pp. 399–408). Durham: Acumen Press.
- Snow, N. E., Wright, J. C., & Warren, M. T. (2019). Virtue Measurement: Theory and Applications. *Ethical Theory and Moral Practice*, 23(2), 277–293. doi:10.1007/s10677-019-10050-6.
- Tanesini, A. (2021). *The Mismeasure of the Self: A Study in Vice Epistemology*. Oxford: Oxford University Press.

T&F Proofs – Not for Distribution

17d Marco Meyer's Response to Commentaries

I am grateful for Alessandra Tanesini pushing me to think more clearly about the metaphysics of virtue underlying my approach, and to Ryan Byerly for challenging me on the methodological claims I made about scale development.

Before I turn to these points, I want to respond to useful challenges mounted by each of them to my definition of social epistemic virtues. I grant right away that my definition of social epistemic virtues needs amending. I defined social epistemic virtues in line with the introductory chapter to this volume as the subset of epistemic virtues aimed at monitoring, adjusting, and ameliorating one's epistemic environment, both for one's own sake and for the sake of others. Viewed together, the comments by Tanesini and Byerly suggest that the definition may be both too broad and too narrow. Moreover, the notion of other-regarding social epistemic virtues needs more explication.

These points are all well-taken. I made no attempt at providing an original definition of social epistemic virtues. Instead, I took my lead from the editors' introduction to the volume. I do think that the notion of social epistemic virtues that they layout gets at aspects of epistemic virtues that are both important and sometimes overlooked. Yet I also agree with Byerly's point that we should not reify social epistemic virtues as distinct from other epistemic virtues. What Byerly suggests is plausible to me: the epistemic virtues that are often discussed in the literature, such as open-mindedness or humility, when possessed in full, may well include many of the social epistemic dimensions that this volume focuses on.

Tanesini contends that the lack of clarity on the nature of other-regarding social epistemic virtues may lead to a challenge for my proposed measurement approach. My approach is mainly focused on self-reports. However, it seems unlikely that people are good judges of how they affect others' epistemic environment. For instance, are people good judges of whether they articulate their knowledge in ways suitable to their audience? I agree that this may well not be so. However, I think that it is worth testing this question empirically. I was highly doubtful that self-reports of intellectual humility would yield adequate self-descriptions.

But the research that establishes construct validity of humility scales convinced me that self-report measures on humility make sense.

Nonetheless, I agree with Tanesini that reports by others may provide important additional data when judging other-regarding social epistemic virtues. The one thing I would caution against is to equate other-regarding social epistemic virtue with success metrics. For instance, we should not measure my ability to articulate my knowledge in an audience-specific way by handing out tests to my students assessing how much they understood from a lecture I gave. That would be to equate a measure of epistemic outcomes with a measure for epistemic virtue. Measures of epistemic virtues and of epistemic outcomes need to be kept distinct if we want to investigate to what extent epistemic virtues drive epistemic outcomes.

Tanesini also urges me to rethink my endorsement of the dispositional account of epistemic virtue that I borrow from Snow. Tanesini's point is that some virtues are not tied to specific virtue-eliciting conditions. To display integrity, for instance, involves consistently behaving according to ethical standards. Being generous partly involves bringing about the very conditions that trigger generous behavior. Tanesini suggests instead to conceptualize epistemic virtues and vices as attitudes. She also notes that the measurement methodology that is typically used when measuring epistemic virtues seems to be consistent with an attitudinal account.

My view is that measures should make as few metaphysical commitments as possible. That way they will be of interest to researchers with a range of views about the nature of epistemic virtue. Finding common ground with the attitudinal approach is therefore very welcome. Rather than endorsing one specific metaphysical conception, we should describe all the different metaphysical views that a suggested measurement approach is compatible with.

Byerly makes helpful points about the extent to which the subtle conceptual distinctions that philosophers are adept at drawing between different epistemic virtues can be measured empirically. I gave a rather pessimistic take. The reason is that the standard methodology in scale development presupposes that subscales within the same survey instrument do not overlap conceptually. This is compatible with the point that Byerly rightly makes, namely that it is common for scales measuring different constructs to nonetheless share several items. I, therefore, agree with Byerly that psychometric methodology does not preclude that we develop overlapping scales, each focused on subtly different epistemic virtues and vices.

One problem with overlapping scales persists. If there is conceptual overlap between two scales, it is very difficult to distinguish empirically whether each scale measures a subtly different epistemic virtue, or whether both scales measure the same virtue, with different levels of success. This difficulty in establishing distinctness of two epistemic

virtues empirically is worrying if you think that positing the existence of an epistemic virtue is warranted only if there is empirical evidence for its existence. By contrast, if one takes the view that being able to draw a conceptual distinction is a sufficient ground to posit the existence of an epistemic virtue, dealing with conceptual overlap is less concerning.

I take the former view. Conceptual distinctions can be misguided or plainly empirically irrelevant. Good measures of epistemic virtues and vices should help us understand which of them matter. Consider personality. There are endless ways of conceptually carving up personality traits. But empirical research has helped us establish a framework of five or six traits that are consistently associated with outcomes. Carving out a trait that captures, say, some aspects of agreeableness and some of open-mindedness is doubtless conceptually possible. No doubt could one develop a scale that measures this trait. But I suspect that it would contribute little to understanding personality.

T&F Proofs – Not for Distribution

18 Learning from ranters

The effect of information resistance on the epistemic quality of social network deliberation

Michael Morreau and Erik J. Olsson

1 Introduction

Consider some matter in which you'd like to know the truth, one way or the other. It could be just about anything. Now, among your friends, colleagues, acquaintances and social media contacts there are many who are open-minded. They learn from relevant evidence, whether from observations and experiments or the testimony of others, and discussing the matter with them is likely to be rewarding. Many people are not open-minded people, though. They are *information resistant* in that they do not appropriately update their beliefs in light of relevant evidence. Engaging with them exposes you to the risk of becoming misinformed because their beliefs not only are fixed but, for all you can tell before the discussion begins, also *false*. What to do?

You might consider limiting discussion to the open-minded people in your network. Anyone unwilling to look, listen and learn then won't get a word in. In general, though, it won't be easy to find out who they are. For reasons that emerge in the next section, just about everybody is information resistant in some matter or another. If there's to be anyone left to talk to, you'll have to identify those who are open-minded with respect to the particular matter that's at stake, and it's not as if people wear such facts about themselves on their sleeves. Another option is simply to include *everybody* in the discussion – even if that risks taking people seriously who do not really deliberate at all: they just say over and again what they think, whether it happens to be true or whether it is false, without learning from anything or anyone. Our main question here is this: how much might unknowingly including these incorrigible sources of misinformation hinder your own open-minded search for the truth?

We approach our question by studying multi-agent computer simulations of Bayesian learning in social networks. The agents in our networks can receive non-social inputs, perhaps from their own observations and the results of experiments, and social inputs in the form of testimony

from other agents. Some of them are open-minded. They learn from both kinds of inputs and consequently, their beliefs change over time. Others agents in our networks are information resistant. Intuitively, they never learn from anything or anyone, and so their beliefs are completely fixed. We focus here on, in particular, *ranters*. These are information-resistant agents that repeatedly broadcast messages conveying their fixed beliefs on the relevant topic. There are among them *true* ranters, whose fixed beliefs are true: they surely are harmless, but there are also *false* ranters whose beliefs are false, and these spread misinformation.

Our findings are somewhat surprising. It turns out that, under seemingly realistic conditions, the presence of false ranters does not much reduce the chance that open-minded agents in the networks will develop high credence in the truth. Sometimes, when deliberation is comparatively short, their presence even increases this chance. Even more surprisingly, perhaps, it's not necessary or even helpful for there to be a balance, with inputs from true ranters as it were making up for misinformation from false ranters in the network, by somehow cancelling them out. Open-minded agents do about as well even when *all* the ranters in their network are false. How is this possible? Scepticism about the reliability of testimony emerges as one important factor. The agents in our networks are able not only to keep track of their sources but also to maintain appropriate levels of trust and distrust in them. Where too much trust is placed in sources initially, before they have demonstrated their reliability, we find that false ranters can quickly lead open-minded enquirers off the track of the truth. This might happen in a real social setting, perhaps, where everybody initially gets the benefit of the doubt, as a matter of respect.

There is much current concern about the consequences of online misinformation from birthers, flat earthers, anti-vaxxers, gaslighters, bullshitters, purveyors of "alternative facts" and false counsellors of all stripes. One way to contain the steady stream of misinformation is censorship, for instance by shutting down the social media accounts of those responsible. This is not without its own risks, though. Censorship can have unintended consequences, such as increasing social and political polarization, when those who identify with the censored views entrench themselves more firmly in their own camps. Censorship furthermore violates social and political norms of inclusiveness and free speech, and can be counterproductive when its targets cry foul and resulting publicity only increases their notoriety, thus amplifying false voices and spreading their misinformation even further than it would otherwise have reached.

Our results bear on the problems of online misinformation. They suggest that, as far anyway as protecting open-minded enquiry is concerned, censorship might often be *unnecessary*. Online interactions lend themselves well to keeping track of sources of information and a careful reckoning of their reliability or otherwise – better, anyway, than

do personal interactions among family and friends. In a range of cases, our results suggest, this monitoring of reputations is, provided there are some sufficiently reliable sources feeding truth into the network, enough to keep the open-minded on the track of the truth. Notice though that our results do *not* show that online misinformation poses any less a threat to well-informed democratic decisions than many people fear. That depends on the proportions. Where too many people prefer the same false option, a few open-minded citizens will not stand a chance no matter how enlightened each one has become. The misinformed will prevail when it comes to a vote.

Our topic is related to the matter of *knowledge* resistance, or irrational failure to accept available knowledge, that has recently been the topic of academic study. We understand this to be the special case of information resistance in which the information to which a proper response is lacking is, indeed, knowledge – or, anyway, something close to knowledge, such as justified belief. Our investigation suggests that social deliberation with knowledge-resistant people need not limit the capacity of other, open-minded people to learn.

In Section 2, we discuss certain mechanisms that can give rise to information resistance, we explain our Bayesian understanding of this topic and briefly mention some related work whose conclusions point in much the same direction as ours. Section 3 explains the basics of *Laputa*, the multi-agent modelling framework that we use to study Bayesian learning in social networks. In Section 4, we discuss the case of a simple social network with just one ranter and one open-minded agent. It illustrates our basic point, and introduces relevant concepts and mechanisms. In Section 5, we argue that the usual way of measuring epistemic value is not suitable for our present study, and reconsider the simple network of Section 4 using a measure that is more suitable. Section 6 takes up a slightly more complex case. We discuss additional simulation results and related work in Section 7 and present our conclusions in Section 8.

2 Preliminaries

Sociologists have identified underlying grounds for information resistance. Mikael Klintman (2019) distinguishes between “Dionysian” tendencies in belief formation, driven by passion and group-centredness, and on the other hand “Apollonian” tendencies or rational and fact-oriented. According to Klintman, knowledge resistance is explained by Dionysian tendencies of group loyalty that encourage us not to deviate from our local culture and its dominant ideology. This phenomenon responds to an evolutionary advantage, namely the fact that a better adaptation to local cultural norms increases the chances of survival and reproduction.¹ Hence, one social ground for information resistance is

that it strengthens bonds within groups and thus enhances collaboration. The more the beliefs of a group deviate from the beliefs of others, the greater the advantage.

Be this as it may, even an Apollonian agent may be information resistant on some issues. One ground for this is belief entrenchment (Quine 1976; Gärdenfors 1988). One might for instance be more responsive to evidence relating to everyday beliefs about changeable matters, such as what there is in the fridge, than to evidence bearing on more basic and deeply entrenched matters such as whether living beings evolved by natural selection, or whether the speed of light is the same in all reference frames. Someone might furthermore be information resistant just because they lack relevant training or other cognitive resources needed to recognize relevant evidence as such and respond appropriately to it. An additional ground for information resistance is that our beliefs are so to speak pinned in place by other propositional attitudes. Someone might believe something is the case, or in any case act as if, and fail to respond appropriately to evidence to the contrary, because of a strong hope that it is the case. Finally, to introduce a technical term, let us say that an agent is *completely decided* with respect to some given proposition if this agent's credence in it is extreme, either 1 or 0. By Bayesian principles of credence updating, an agent that is completely decided about some proposition won't ever change her credence in it, and is information resistant.

Now we can put our question in more-precise terms. Suppose an open-minded agent wishes to know the truth in some given matter, expressed by some proposition, p . Let it be you, the reader. There are perhaps other open-minded people in your social network who share this wish. There might also be some though who are information-resistant with respect to p . They do not respond appropriately to evidence bearing on the truth of p from observations, experiments and social interactions with others; and initially, anyway, before you begin deliberating together with them, you cannot tell who in your network is open-minded and who is not. From behind this veil of ignorance, you have these two options (among others): You can enter into discussion with everyone in the network, treating them all the same way. Or else you can put your fingers in your ears, figuratively speaking, and only trust your own observations, experiments or other non-social sources of information. Which is the best option?

We approach our topic from a Bayesian perspective, using the method of multi-agent modelling. We use in particular the social simulation framework *Laputa* (Olsson 2011) to model both resistance to information from the non-social world and resistance to social information from peers. The information-resistant agents we're interested in here are what we call *ranters*. These are agents who are resistant both to non-social and to social evidence concerning p , whose credence in p is sufficiently high or low so that they are willing to assert either that p or that *not* p ,

as the case may be, and who repeatedly do express their fixed belief concerning this matter within the network. Intuitively, a ranter is someone who tries to convince others of his view while not being himself able or willing to adjust to new evidence of any kind. The next section has further details on the Laputa framework and the use we make of it.

3 The Laputa framework for Bayesian social epistemology

Laputa enables us to study networks of agents, intuitively enquirers.² An agent's belief state is represented by a probability distribution corresponding to their degree of belief in some proposition in question, again call it p . This proposition is assumed to be true. Degrees of belief (or credence) are updated by conditionalization on the evidence. This evidence either comes from a non-social source (perhaps observations or experiments), or else it comes in the form of a message from another agent in the network. These messages sent and received are of two kinds: p or not- p . Thus, Laputa models network activity in response to a binary issue. At any step in a deliberation, agents can communicate with other agents in the network, or they can receive information from their non-social source. Distributions determining the chance of communication, of receiving outside information and so on, at any given point in the deliberation, are parameters in the model.

Suppose a source, S , sends the message p . One important point is that the evidence a receiver gains from this message is not just p itself. It is that S reported that p . Similarly, *mutatis mutandis*, when a source sends the message not- p . This is important because it opens up the possibility of not taking testimony at face value. The Laputa framework incorporates a Bayesian mechanism for representing the degree to which an agent trusts her own enquiry (that is, her nonsocial or "outside" source) as well as the extent to which she trusts the different agents in the network. Trust here means perceived reliability and is represented as a "trust function" over all possible reliability profiles – from being systematically truth-telling to being systematically false-telling – representing how likely those profiles are taken to be at a given stage in the deliberation. For some purposes, trust can be represented by a single number: the expected value of the trust function. An enquirer's new trust function after receiving information is obtained via conditionalization on the evidence. In the simple case in which the enquirer has a normally distributed trust function with an expected value of 0.5 and assigns p a degree of belief exceeding 0.5, the enquirer will, upon receiving repeated confirming messages from one source, update her trust function so that it approaches a function having expected value 1, representing full trust in the source. Interestingly, representing trust by a function rather than a single number allows for complex interactions between different parameters. Two agents who assign

the same degree of belief to p , who have trust functions with the same expected value, and who receive exactly the same information (say, from enquiry) may, nevertheless, depending on their initial trust functions, end up with very different degrees of belief and trust functions.

A computer implementation automatically computes changes in credence and trust. This greatly facilitates investigation into the model and its consequences (see Olsson, 2011, for an overview). Consider the “Laputa table” (Table 18.1) for updating belief and trust (see Olsson, 2013, for derivations).

Table 18.1 summarizes how updating in Laputa works. Consider, for example, the upper left-most cell. This deals with the case in which an agent receives from a trusted source a message that is *expected*, in the sense that the content of this message, say p , is already assigned by the receiving agent a higher credence than 0.5. That the source is trusted means that the receiving agent assigns a trust function to the source such that the expected value of that function is higher than 0.5. What should happen in this case? The plus sign here means that the receiving agent will strengthen her current belief. In our example, it means that she will believe even more strongly that p is the case. The up-arrow means that the receiving agent will trust the source even more. Similarly, the minus sign in Table 18.1 means that the receiving agent weakens her current belief, and the down-arrow means that she trusts the source less than she did before. For example, agents downgrade their trust in a trusted source when presented with an unexpected message from that source. It is important to understand that the rules described in Table 18.1 are derived rules in the sense that they follow from the underlying Bayesian machinery. They are not separate stipulations.³

Following Goldman (1999), epistemic value in Laputa is equated with veritistic value. The main idea is that an agent ideally has full belief in the truth. Thus a credence of 1 in the truth yields maximum veritistic value. More generally, the closer you are to assigning credence 1 in the truth, the better. Thus, we can define the veritistic value of assigning credence X in the truth, simply, as X . We will prefer the term “epistemic value”, abbreviated E-value. In the case of a group of agents, the epistemic value is the average epistemic value of the members of the group. In Laputa, we are interested in the effect of deliberation on the epistemic

Table 18.1 Derived updating rules for belief (credence) and trust

	<i>Message expected</i>	<i>Neither nor</i>	<i>Message unexpected</i>
Source trusted	+ (↑)	↑ ()	- (↓)
Neither nor	0 (↑)	0 ()	0 (↓)
Source distrusted	- (↑)	↓ ()	+ (↓)

value of the group, that is in the difference between the initial E-value and the final E-value. This is abbreviated Δ E-value.

4 Learning from one ranter

Consider first a simple network with only two agents (Figure 18.1). It illustrates the Laputa framework and the mechanism underlying results to come.

One of the agents is a false ranter. Technically, this agent is completely decided and wrong. He assigns an extreme credence of 0 to the proposition p (the truth). He does not perform any enquiry – the “enquiry chance” is set to 0. But even if he did, since he is a Bayesian agent with an extreme credence, it wouldn’t make any difference. The other agent is *completely undecided* about p , having credence 0.5. Moreover, this agent is open-minded. She listens to her outside source, which represents in our model personal observations and experiments. We assume that the probability that she does so at any given step in the enquiry is 0.6 and that the reliability of this source is 0.7: it gives the right answer about 7 times out of 10. Initially, she has some modest degree of trust in her outside source. She’s entitled to it perhaps because the quality of the source is something that’s under her control: she’s the one who makes the observations or set up the experiments. We model this by drawing the open-minded agent’s initial trust in the outside source from a uniform distribution centred at 0.6. That is, in any given deliberation, her trust in this source starts off at about 0.6, sometimes slightly lower or higher.



Figure 18.1 An open-minded, undecided and somewhat competent agent prepared to listen to a ranter who, unbeknownst to her, is on the wrong side of the debate.

The ranter is information resistant and doesn't listen to the open-minded agent: there is no communication link leading towards him. But even if he did listen, since he is a Bayesian agent with an extreme credence, it wouldn't make any difference. The open-minded agent, on the other hand, being open-minded, does listen to the ranter who, at any given time, we assume, sends a message with a probability of 0.6. Initially, the open-minded agent has no idea whatsoever how much to trust these messages. We represent this by having her draw her initial degree of trust in him from a uniform distribution centred at 0.5. This means that her initial trust in him will be around 0.5, perhaps slightly lower or slightly higher. Notice the difference in her initial trust in her two kinds of sources. She starts off with some degree of trust in her own observations and experiments, but she is completely uncommitted when it comes both to her trust in the other agent and her belief in the proposition p that is the subject of the enquiry.

The course of a typical run of this network is pictured in Figure 18.2. The open-minded agent becomes increasingly sure that the other agent is systematically unreliable, and starts treating what he says as evidence to the contrary.

Figure 18.2 shows how the open-minded agent's credence in p quickly converges to full belief in p in spite of the fact that the ranter is always telling her that not- p . Her trust in the ranter slowly but surely erodes, starting with 0.5 and ending up below 0.4 after 20 simulation steps. Meanwhile, her trust in her own enquiry increases somewhat but not markedly, approaching the actual reliability of her inquiry (0.7). In the process, the E-value improves quite significantly.

Using Laputa's batch simulation feature, we can study the epistemic performance of this small network over a large number of simulations. In this case, we let Laputa run the network 10,000 times, each time for 30 steps. The average epistemic value over all these runs was 0.1972 (± 0.003 with 95% confidence). This means that the average improvement of epistemic value for the agents in the network was 0.1972. Of course, all this improvement comes from the open-minded agent because the ranter, being unresponsive to evidence, does not change his credence in p . (Some people never learn.)

One might suppose that this effect is solely due to the open-minded monitoring and updating of trust. However, even if we turn off social-trust updating, there is still an improvement in the average epistemic value for the network (although it is on the average somewhat lower for 30-step simulations). The improvement is 0.1881 (± 0.002 with 95% confidence).⁴ Increasing the number of steps to 50 yields an improvement of 0.2237. What drives the effect is that the trust in the somewhat reliable outside source is greater than the trust in the ranter.

Of course, in terms of average epistemic value over all agents in the network, we get higher epistemic performance if we leave the open-minded

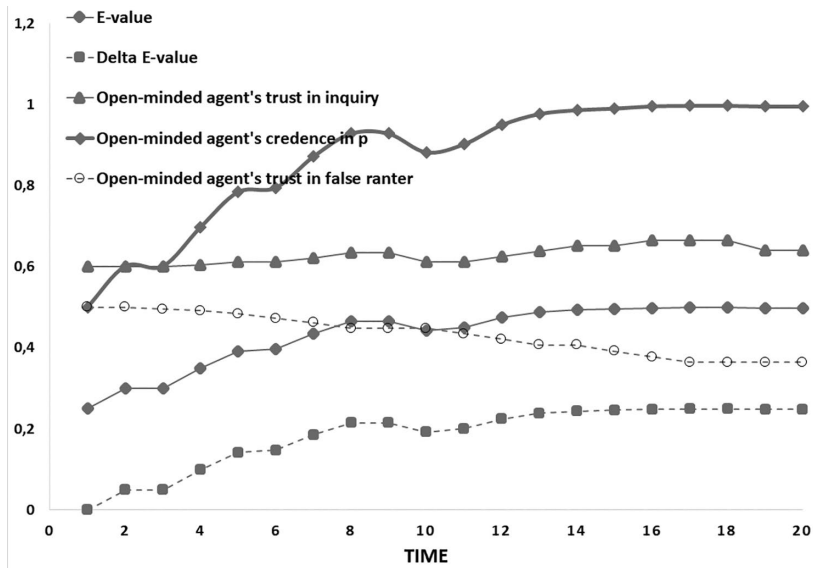


Figure 18.2 Result of a typical run of the two-person network in Figure 18.1.

agent to her own devices. Thus, if we remove the ranter from the network, her average gain over 10,000 simulations, each 30 steps long, is 0.3781, compared to 0.1972 with all the false ranting. From this perspective, the ranter is an unwelcome distraction, whether trust is updated or not. The next section proposes a more nuanced way of looking at epistemic value in situations like the present one. It gives a very different perspective on this case.

Trust updating often makes the network stabilize sooner because it enables the open-minded agent to identify the closed-minded agent as a false ranter and eventually even learn something from him, i.e. use this insight in updating her own credence in p . However, it is a double-edged sword. Consider turning off enquiry trust in the case of the single open-minded enquirer. The average epistemic value gain is then 0.3900, which is even better than the 0.3781 we got when enquiry trust was updated (again 10,000 simulations, each lasting 30 steps). While updating trust is sometimes good because it speeds up a process that is already on the right track, it is sometimes bad because it can get you off on the wrong track. Although the outside source in our simulations is 70% reliable, this also means that it delivers the wrong result in 30% of the cases. Thus, it can happen that the first, say, ten results are all “bad”: the outside source reports not- p when in fact p is true (cf. Vallinder and Olsson 2013a). If trust is dynamically updated depending on the proposition reported, this may suggest that enquiry is in fact systematically unreliable.

When the “good” results start coming in, which they will, in the end, they will be viewed as evidence to the contrary, meaning that the agent has painted herself into an epistemic corner or, using another metaphor, entered a spiral of distrust.

5 Epistemic value – a more general perspective

Laputa’s built-in approach to measuring epistemic value at the network level is to average epistemic value over all the agents in a network. This follows the account of veritistic value in Goldman (1999). We propose now to measure epistemic value by averaging over some group of the agents that, for whatever reason, are of special interest. The built-in measure, of course, is just the limiting case in which the relevant group includes all the agents.

One use for the general measure is in assessing your own epistemic prospects from behind a veil of ignorance. Suppose you’re deciding whether to enter into deliberation about some proposition within a social network. There is much that you know about your own situation: your own credence in this proposition, perhaps; and that you yourself are open-minded, being willing to learn from observations, experiments and the testimony of others. But there are other relevant things you don’t know, such as what the other agents’ credences are, whether they are open-minded, the reliability of their sources, who’s in their networks, and so on. The epistemic benefit you expect from engaging can be assessed by averaging just over the group of those agents whose epistemic situations agree with what you know about your own situation, but which differ from one to the next in precisely those aspects of your own situation about which you are unsure. These are the agents whose epistemic situation might, as far as you know, be your own.

Focusing on the learning of some subgroup of agents is useful in our study. One group that is of particular interest is the group of all open-minded agents. Averaging just over them, we have what we call *Epistemic Value for the Open-Minded*, or EVO. It is easily computed from the epistemic value as calculated by Laputa’s built-in metric: we multiply by the total number of agents in the network, to undo the averaging over them all, and then divide the result by the number of open-minded agents.

EVO gives a better picture of learning in our networks because it restricts attention to only the agents who are learning. To see how this makes a difference, we return now to the simple network of Section 4. The increase in epistemic value, as measured using the built-in metric, was 0.1972 (10,000 simulations, 30 steps). Computing the EVO gain instead gives us $2 \times 0.1972 = 0.3944$. This is interesting because the open-minded agent’s gain on her own, without the ranter, is a full two percentage points lower, just 0.3781. We see from this perspective that

the open-minded enquirer actually *benefitted* from having the ranter in her network! She did even better than she would have done on her own and turning off the dynamics of enquiry trust, in which case the epistemic gain, as we saw, would be 0.3900.

6 The case of two ranters

Next, we investigate what happens with an open-minded agent listening not to one but two ranters – one firmly believing p (this is a true ranter) and the other one firmly believing $not-p$ (a false ranter) (Figure 18.3). Otherwise, this scenario is the same as in the previous one-ranter scenario.

The course of a typical simulation trial is shown in Figure 18.4.

Again, after an initial misfortune whereby (as the detailed logs show) the open-minded agent received a false result from enquiry, her credence in p converges to 1, that is full belief in the truth. Meanwhile, her trust in the false ranter decreases steadily whereas her trust in the true ranter increases. Since the open-minded agent's credence approaches full belief in the truth, there is a distinct epistemic improvement.

The results of our batch simulations (averages over 10,000 trials) are summarized in Figure 18.5, where the x-axis is the number of steps in the simulations. There is, as before, a small error in the third decimal when computing the difference in E-value (95% confidence level). The difference in EVO results from multiplying the difference in E-value by 3 (= the number of enquirers in the network). In the same diagram, we plot also simulation results with the two ranters removed.

Note that in the situation where the undecided agent is left to her own devices, Δ E-value and Δ EVO coincide. We see that, as expected, the development of Δ E-value is not very impressive in the two-ranter scenario (bottom curve in Figure 18.5) in comparison with that of the single-agent network (middle curve in Figure 18.5). However, things look very different using the EVO metric. Then the two-ranter situation

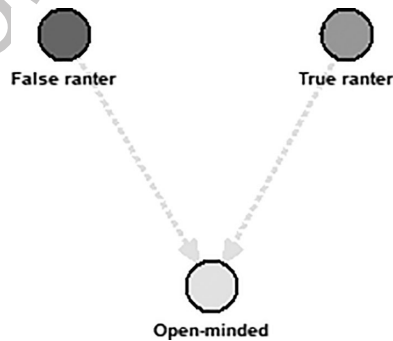


Figure 18.3 Listening to ranters on opposite sides of the issue.

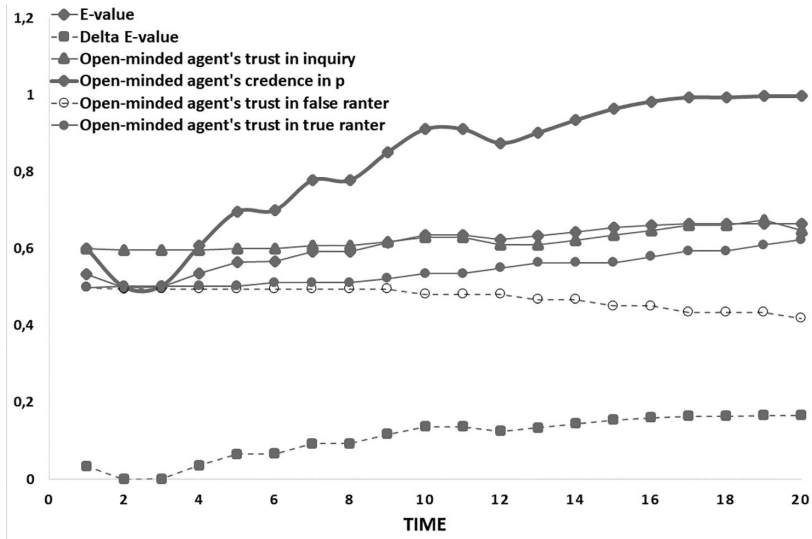


Figure 18.4 Results of a typical trial in the case with two ranners on opposite sides.

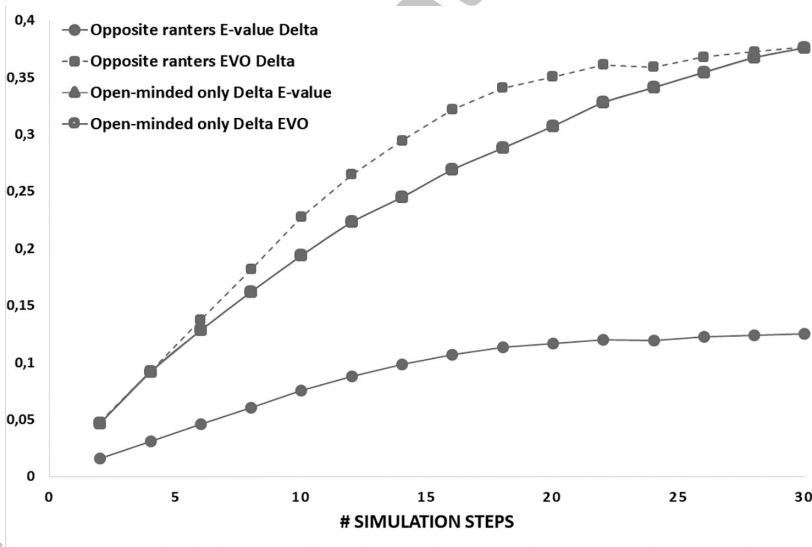


Figure 18.5 Simulation results for the network with two ranners on opposite sides and with the ranners removed (see Figure 18.6).

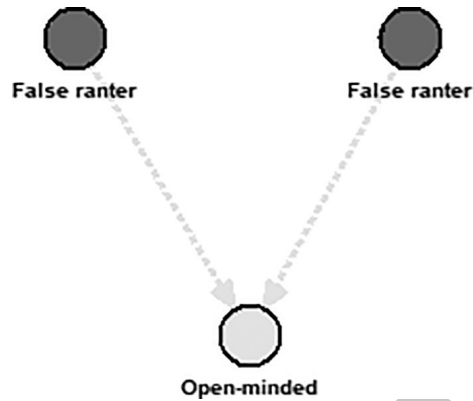


Figure 18.6 Listening to two false ranters.

even outperforms the single-agent scenario in simulations with a moderate number of steps (5–30). For shorter or longer simulations, the two scenarios give rise to the same Δ EVO.

Surprisingly, our conclusions remain essentially the same even if both ranters are on the wrong side of the issue, as the following example illustrates (Figure 18.6):

Figure 18.7 shows the results of batch simulations for the network in Figure 18.6 (average over 10,000 trials). There we have also plotted the same results for the case where the open-minded agent has disabled social trust updating.

First, the top curve in Figure 18.7, depicting the Δ EVO as a function of the number of simulation steps for the two false ranters case, is practically identical to that for the case of two ranters on opposite sides (top curve in Figure 18.5), with updating of social trust enabled. Disabling social trust gives rise to less pronounced improvement in epistemic value, for a moderate number of simulation steps, if epistemic value is measured by EVO (second curve from the top). The same is true if epistemic value is measured by E-value (second vs. first curve from the bottom in Figure 18.7).

7 Discussion

So far, we have been looking at networks of only two or three people. This was sufficient to establish our main point: that there are situations in which ranting does not affect the epistemic value of deliberation or even is epistemically beneficial, from the perspective of

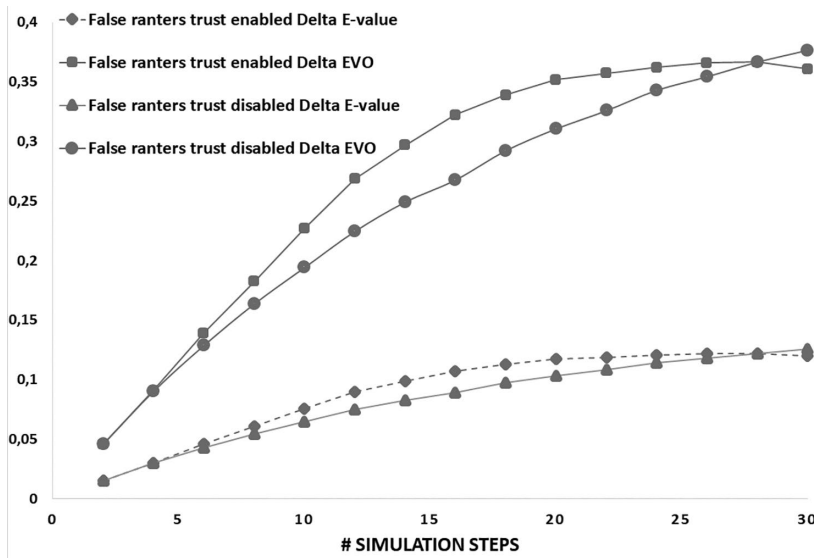


Figure 18.7 Simulation results for the network with two false ranters (see Figure 18.6).

initially undecided agents. One might wonder what happens in larger networks. To find out, we looked at a network representing a board in an organization. We limited ourselves to relatively long simulations (30 steps).

The simulated board has ten members. Since in a board everyone talks and listens to everyone else the network is fully connected, with arrows from each agent to every other agent (Figure 18.8). We modelled several cases, from all agents being open-minded and initially undecided, to all but two being ranters, with an equal proportion of ranters on either side of the issue. Figure 18.8 depicts a case of four ranters.

The results for Δ E-value and Δ EVO are summarized in Table 18.2 (30 steps, average over 10,000 simulation trials).

What we expected, based on earlier results, was that the E-value would decrease with the number of ranters, but that the epistemic value for the initially undecided agents, captured by EVO, would remain essentially the same. As shown in Table 18.2, this is indeed what we observed. The benefit of deliberation in terms of E-value decreases, as it must, when additional non-learning ranters pull the overall average down, but the benefit of deliberation in terms of EVO remains about the same, with just some small variation within the margin of error. So, once more, including ranters does not diminish the epistemic value of deliberation for open-minded agents in longer simulations.

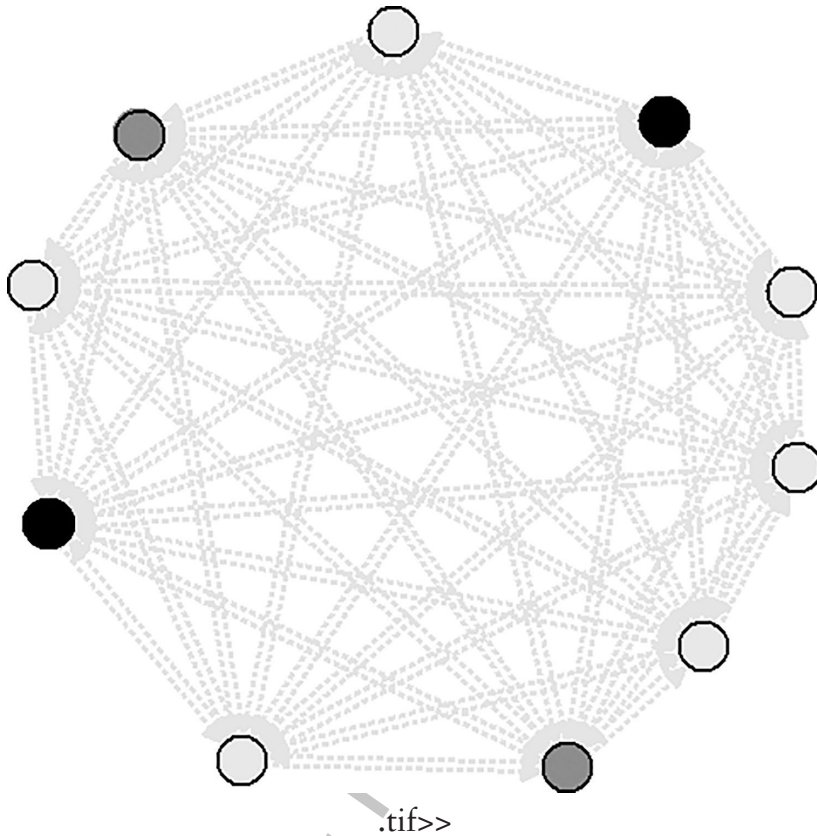


Figure 18.8 A board with four ranters, two on each side of the issue.

Table 18.2 Epistemic value for the board with different proportions of ranters (30 steps, average over 10,000 simulation trials)

	$\Delta E\text{-value}$	ΔEVO
All open-minded	0.1691	0.1691
8 open-minded, 2 ranters	0.1309	0.1636
6 open-minded, 4 ranters	0.0950	0.1583
4 open-minded, 6 ranters	0.0640	0.16
2 open-minded, 8 ranters	0.0312	0.156

While we are not aware of any study of knowledge resistance *per se* there are related studies suggesting (a) that restricting social information, or equivalently, being resistant to such information can sometimes be epistemically beneficial and (b) that it can be similarly beneficial to be somewhat resistant to belief-contravening information.

Zollman (2007) studied the effect of reducing the density of a social network, i.e. the number of communicative links between the agents on the epistemic performance of the network, using a Bayesian model suggested by economists Bala and Goyal (1998). The agents in the network were thought of as scientists, each conducting an experiment and transmitting the result to their network peers. The surprising result was that in “some contexts, a community of scientists is, as a whole, more reliable when its members are less aware of their colleagues’ experimental results” due to a sparser network structure (p. 574). However, denser networks exhibited quicker convergence on a stable group opinion. A similar “less is more” effect was found in Hahn et al. (2018) using the present Laputa Bayesian setting.⁵ They studied a wide range of network structures and provided a detailed statistical analysis concerning the exact contribution of various network metrics to collective competence. Specifically, they found that 96% of the variation in collective competence across networks could be attributed to differences in amount of connectivity (density) and clustering, which were both found to be negatively correlated with collective competence. A study of bandwagon or “group think” effects indicated that both connectivity and clustering increase the probability that the network, wholly or partly, locks into a false group opinion.

These studies suggest that listening and talking *less* to your friends or colleagues can sometimes be epistemically beneficial for the collective. The limiting case would be a group of agents who *never* listen or talk to anyone, and therefore qualify as socially information resistant. Based on these studies one would expect that at least sometimes such a group would be better off than networked groups. However, though related, this is not the matter that we have addressed here. We have investigated the effects on social deliberation of agents who don’t respond to any kind of inputs, whether from friends and colleagues or from observations or experiments, but who even so broadcast their fixed opinions to everyone listening.

Another related study is Vallinder and Olsson (2013a), which addresses the epistemic value of overconfidence also using Laputa. Many psychological studies have found that people think they are more reliable than they actually are (e.g. Harvey, 1997; Johnson and Fowler 2011). Using computer simulation in a Bayesian setting, Vallinder and Olsson showed that agents are indeed sometimes epistemically better off collectively overestimating the reliability of their own enquiry (network-external source). They also found that people rarely are better off overestimating the reliability of others. On the basis of these findings, they suggested that overconfidence in enquiry may be valuable because it makes agents less vulnerable to strings of “bad” results, in a context in which trust is updated dynamically. Dynamically updated trust in enquiry may otherwise, as we saw, lead more easily to a vicious spiral whereby the agents

increasingly distrust the input they get from the external world, taking it rather as “evidence to the contrary”. Overconfidence clearly is related to information resistance: an agent who is overconfident vis-à-vis enquiry is more resistant to unexpected information coming from enquiry; such information will have less impact on the agent’s credences. While Vallinder and Olsson’s study investigated a distinct, if related, phenomenon, it does suggest that the epistemic consequences of information resistance might not all be bad.

8 Conclusion

The question we have addressed is how, if at all, the presence of information-resistant agents influences the epistemic value of group deliberation. We approached this question from a Bayesian perspective, explicating information resistance as failure to appropriately update credence in light of incoming information. In particular, we focused on “ranters”: information-resistant agents that obstinately communicate their fixed opinion in their networked group. In evaluating the epistemic effect of including ranters in a network of open-minded agents, we found it instructive to measure not the (average) epistemic value for the network as a whole, but rather the (average) epistemic value for just the open-minded agents.

Our study suggests that including ranters has little or no negative effect on the epistemic value of social deliberation. Including them can even be epistemically beneficial if the open-minded agents in the network continuously update their trust or distrust in other agents. This dynamics of trust has in the networks we studied the effect that open-minded enquirers come to treat false ranters – quite mistakenly, but still fortuitously – as if they were genuine though, so to speak, “upside-down” sources of information. Their messages are treated as evidence to the contrary, though in fact like stopped clocks they’re not tracking the truth at all.⁶

What makes possible this happy inversion of consistently false testimony seems to be that the open-minded agents have somewhat reliable non-social or “external” sources of information, in which they have furthermore at least some initial trust. Consulting these external sources more often than not injects truth into the enquiry, nudging credences in the right direction. False ranters might initially disturb this open-minded progress towards the truth; but, provided trust in them is not initially too high, before long the external sources reveal their persistent misinformation for what it is. Their false ranting eventually catches up on them. They develop reputations for speaking falsely, and their poor reputations make their ranting harmless.

Excluding information-resistant people from debates requires recognizing them as such, which realistically must consume resources. Censorship, de-platforming and other ways of excluding them furthermore

violate social and democratic values of inclusiveness and free speech. Our study suggests that, in some contexts anyway, the epistemic benefits of excluding false ranters are not worth these costs. Where it is feasible to keep tabs on developing reputations for speaking the truth, and for speaking falsely, and to count or discount people's contributions accordingly, it might be better simply to let everyone have their say. This close monitoring of sources and reputations might often be feasible for instance in the case of online exchanges of information. Our results are preliminary. Establishing them more firmly will require further modelling, empirical calibration, and simulation work.

Notes

- 1 Another, complementary, explanation centres on the influence on our beliefs of "negativity bias", an evolutionary tendency to pay more attention to negative information to increase our chances of survival, on our beliefs. Negativity bias makes us pay more attention to the risks of certain behaviours, and resist knowledge about their associated benefits. Cf. Costa-Font (2020).
- 2 The Laputa model has been applied to a number of other problems in epistemology, such as norms of assertion (Olsson and Vallinder 2013a), the argument from disagreement (Vallinder and Olsson 2013b), the problem of jury size in law (Angere et al. 2015) and peer disagreement (Olsson 2018).
- 3 Collins et al. (2018) examined, theoretically and empirically, the implications of using, in the spirit of Laputa, message content as a cue to source reliability. They presented a set of experiments examining the relationship between source information and message content in people's responses to simple communications. The results showed that people spontaneously revise their beliefs in the reliability of the source on the basis of the expectedness of a source's claim and, conversely, adjust message impact by perceived reliability, much as updating works in Laputa. Specifically, people were happy downgrading their trust in a source when presented with an unexpected message from that source.
- 4 The accuracy of all simulations results to come is of the same magnitude, that is, the results allow for a small error in the third decimal.
- 5 For a related study, see Angere and Olsson (2017).
- 6 This epistemic boost based on the dynamics of trust and distrust, however, exists only for deliberations of moderate length (we observed it in simulations of 5–30 steps). For shorter and longer deliberations, the epistemic effect is the same whether or not the undecided agents update their trust in the ranters in light of their communications.

References

- Angere, S., and Olsson, E. J. (2017). Publish late, publish rarely! Network density and group performance in scientific communication. In Boyer, T., Mayo-Wilson, C., & Weisberg, M. (Eds.), *Scientific Collaboration and Collective Knowledge*. Oxford University Press: 34–62.
- Angere, S., Olsson, E. J., and Genot, E. (2015). Inquiry and deliberation in judicial systems: the problem of jury size. In Baskent, C. (Ed.), *Perspectives on Interrogative Models of Inquiry: Developments in Inquiry and Questions*. Springer: Dordrecht.

- Bala, V., and Goyal, S. (1998). Learning from neighbours, *Review of Economic Studies* 65: 565–621.
- Collins, P. J., Hahn, U., von Gerber, Y., and Olsson, E. J. (2018). The bi-directional relationship between source characteristics and message content, *Frontiers in Psychology* 9. <https://doi.org/10.3389/fpsyg.2018.00018>
- Costa-Font, J. (2020). Review of knowledge resistance, *Journal of Behavioral Economics* 86. Published online. <https://doi.org/10.1016/j.socec.2020.101540>
- Hahn, U., Hansen, J. U., and Olsson, E. J. (2018). Truth tracking performance of social networks: How connectivity and clustering can make groups less competent, *Synthese*: 1–31. <https://link.springer.com/article/10.1007/s11229-018-01936-6>
- Gärdenfors, P. (1988). *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge, MA, London: MIT Press.
- Goldman, A. I. (1999). *Knowledge in a Social World*. New York: Oxford University Press.
- Harvey, N. (1997). Confidence in judgment, *Trends in Cognitive Science* 1(2): 78–82.
- Johnson, D., and Fowler J. (2011). The evolution of overconfidence, *Nature* 477: 317–320.
- Klintman, M. (2019). *Knowledge Resistance: How We Avoid Insights from Others*. Manchester: Manchester University Press.
- Olsson, E. J. (2011). A simulation approach to veritistic social epistemology, *Episteme* 8(2): 127–143.
- Olsson, E. J. (2013). A Bayesian simulation model of group deliberation and polarization. In Zenker, F. (Ed.), *Bayesian Argumentation: The Practical Side of Probability*, Synthese Library, New York: Springer: 113–134.
- Olsson, E. J. (2018). A diachronic perspective on peer disagreement in veritistic social epistemology, *Synthese*: 1–19. <https://link.springer.com/article/10.1007%2Fs11229-018-01935-7>
- Olsson, E. J. (forthcoming). Why Bayesian agents polarize. In Broncano-Berrolca, F., & Carter, A. (Eds.), *The Epistemology of Group Disagreement*. Routledge.
- Olsson, E. J., and Vallinder, A. (2013). Norms of assertion and communication in social networks, *Synthese* 190: 1437–1454.
- Quine, W. V. O. (1976). Two Dogmas of Empiricism. In: Harding S. G. (Ed.), *Can Theories Be Refuted?* Synthese Library (Monographs on Epistemology, Logic, Methodology, Philosophy of Science, Sociology of Science and of Knowledge, and on the Mathematical Methods of Social and Behavioral Sciences), Vol. 81. Springer, Dordrecht.
- Vallinder, A., and Olsson, E. J. (2013a). Trust and the value of overconfidence: A Bayesian perspective on social network communication, *Synthese* 191(9): 1991–2007.
- Vallinder, A., and Olsson, E. J. (2013b). Do computer simulations support the argument from disagreement? *Synthese* 190: 1437–1454.
- Zollman, K. S. (2007). The communication structure of epistemic communities, *Philosophy of Science* 74: 574–587.

18b Commentary from Georgi Gardiner

Antisocial Modelling

Morrau and Olsson employ a simulation to investigate epistemic effects of false assertions. Ranters are defined as “information-resistant agents that repeatedly broadcast messages conveying their fixed beliefs on the relevant topic”, where “information-resistant” means they never update beliefs in light of new evidence. There are two kinds: False ranters only ever convey false claims; true ranters only ever convey true claims. “Open-minded” agents, by contrast, update credences in light of evidence.

Morrau and Olsson conclude

Our study suggests that including ranters has little or no negative effect on the epistemic value of social deliberation. Including them can even be epistemically beneficial if the open-minded agents in the network continuously update their trust or distrust in other agents.

Even if 80% of agents in the deliberation are false ranters, they claim, “the benefit of deliberation in terms of [their measure of epistemic value] remains about the same ... So, [...] including ranters does not diminish the epistemic value of deliberation for open-minded agents”. Call this the “surprising result”.

This result relies on open-minded agents’ ability to track who speaks reliably and to “trust update”, that is, to adjust the degree to which others’ assertions affect their credences. Open-minded agents must even begin to treat false ranters’ assertions that p as evidence that not p . The surprising result also relies on their novel interpretation of epistemic value. By “epistemic value” Morreau and Olsson mean “epistemic value for the open-minded” (EVO). EVO measures the average degree of divergence between credence and truth-value, when restricted to open-minded agents. It excludes ranters’ credences.

In what follows, I first sketch a concern about the realism of the simulation. I then enumerate some candidates for epistemic value not reflected in EVO and, in some cases, not reflected in the simulation more generally.

In the simulation, traits are binary, uniform, and unchanging. Everyone either Bayesian updates flawlessly on all evidence or they never alter any beliefs at all. The latter – ranters – have either comprehensively true or comprehensively false beliefs. All non-social evidence sources emit uniform chance of accuracy, namely 70%. Reality is more complex: A person speaks truly sometimes, and not others, and are more reliable on some topics. People might aptly modulate expressions of confidence, but lapse into overconfidence when discussing politics, or become too diffident around the highly educated. Real-life testimony usually consists of multiple interrelated claims and so can be partially accurate. Non-social sources are also heterogeneous. Some kinds are more reliable than others. And both inquirers and their sources change – we can become better at interpreting and assessing evidence. This bears on whether the simulation reflects real-life. It is harder to discern, retain, and employ track records when they are complex, evolving, and nuanced, and when we lack dependable or predictable external sources to calibrate against. This variability makes trust updating – an essential component for their result – more onerous and prone to error.

Secondly, the epistemic value provided by false ranters requires that open-minded agents treat the false ranters' assertion that p as evidence that not p . This mechanism underwrites the “surprising result”. But I doubt people do this. It seems contrary to interpersonal interaction and inimical to the institution of testimony. Even considering people I deem most epistemically irresponsible, such as children or anti-vaccine astrolgists, I still don't do it. If they assert some arbitrary claim, I consider it evidence or, at worst, perhaps neutral. I do not consider it evidence *against* p .

There are special cases where assertions can be a reason to believe the opposite. But they are marginal, require significant background evidence and context, and are about limited domains of assertion. If your acquaintance has contrasting political views, for example, their assertion about the best political candidate can be counterevidence. Tribal, polarized political landscapes can also lead to people taking assertions as counterevidence or, at least, purporting to. Examples include assertions like “Mask-wearing is a safe, effective way to block COVID transmission”. Perhaps we also treat some kinds of aesthetic judgement as counterevidence. We might learn that a person's musical taste differs so radically from ours, for example, that their liking a musician is evidence that we will not, even if we know nothing else about the musician, such as the genre. But such examples are constrained to limited domains, and do not apply to all assertions.¹

These two concerns about realism intertwine. We do not treat assertions as counterevidence because real people are not close to uniformly unreliable, like false ranters in the simulation. Perhaps if false ranters lived among us, we would do this. But given realistic heterogeneous

track records, I doubt we realistically can “treat what he says as evidence to the contrary” for all or arbitrary-selected assertions of another person. Simulations never aim to perfectly mirror reality, of course, but Morreau and Olsson’s idealizations threaten their conclusion and its applicability to society.

I now turn to costs that Morreau and Olsson do not countenance. I present a miscellany of candidate values and disvalues. Some are measurable in their existing simulations or can be incorporated into new simulations. Others not. Some values accord with veritistic approaches; others do not. And, finally, theorists disagree about which candidates are indeed valuable, and whether that value is epistemic. I lack space to investigate these categorizations and questions; I simply enumerate some candidate values.

Morreau and Olsson’s measure of epistemic value, EVO, concerns the average accuracy of open-minded agents in the network after a set number of interactions. EVO disregards ranters’ credences. Morreau and Olsson celebrate that EVO remains high even when 80% of group members are false ranters: “[False ranters] develop reputations for speaking falsely, [which makes] their ranting harmless” because they are not believed. And, since their assertions that p are treated as evidence that not p , “including them can even be epistemically beneficial”.

Morreau and Olsson’s justification for excluding ranters from their measure of epistemic value is twofold. Firstly, ranters don’t learn, so – they claim – their errors are irrelevant to a study of social learning. Morreau and Olsson comment “some people never learn”. But, taken literally, this is false: All people learn. Indeed, since the simulation’s ranters never adjust “beliefs” in light of evidence, and every “belief” is false, I am doubtful the nodes warrant the labels “agent” and “belief”, as opposed to mere information sources. If impervious non-learning justifies excluding them from the measure of epistemic value, it may correspondingly cast doubt on the simulation’s claim to model epistemic interactions amongst different kinds of agents: Ranters aren’t agents.

On the other hand, if ranters are agents, excluding them risks qualifying as creative accountancy: Poorly performing agents are omitted from the record. Discounting them might be elitist, condemning them as deplorables. An epistemic community is better if it improves the epistemic character and conduct of its members, and society should discourage evidence resistance. In real life, the presence of other delinquents reinforces and sustains epistemic delinquency. A lower incidence of ranting might help rehabilitate ranters. Even setting aside developmental, diachronic considerations, and focusing only on graded alethic accuracy after a number of simulation steps, distribution of accuracy can matter. Non-polarized groups are plausibly better. EVO, as a measure of epistemic value, overlooks these potential disvalues.

Secondly, Morreau and Olsson claim that excluding ranters from the measure of epistemic value better addresses their research question, namely, “how much might unknowingly including [...] incorrigible sources of misinformation hinder your own open-minded search for the truth?” By excluding ranters, EVO better approximates expected epistemic value for “the agents whose epistemic situation might, as far as you know, be your own”.

In response, firstly, open-minded people suffer when surrounded by incorrigible ranters. This chapter sketches some costs they encounter. So from the perspective of open-minded agents, EVO overlooks relevant values. And, secondly, epistemology should also consider an impartial perspective. It should ask not only whether the community is propitious “for me”, but also how the epistemic community is faring overall. If 80% of members have 100% false beliefs, the answer is *badly*.

An epistemic community is better if most beliefs and assertions are true. In the simulations, many assertions are false, disbelieved, or treated as counterevidence. A prevalence of true assertions has instrumental epistemic value. Discerning and recalling who is unreliable demands cognitive resources, which has opportunity costs. Trust updating requires dynamic accountancy. One must track whose previous assertions conflicted with one’s antecedent beliefs and weight their future testimony accordingly. The effort is better invested elsewhere. And trust updating is fallible. Learning is easier if we can generally believe people.

Indeed, absent background knowledge about the bifurcated epistemic community – that is, ranters and others – encountering false ranters in the simulation should decrease trust in testimony simpliciter. Testimony appears – and is – unreliable.

A prevalence of true assertions may also exhibit non-instrumental epistemic value. There is plausibly cognitive value in attention being directed towards the right things. Perhaps proper attentional patterns are constitutive of flourishing, for example.² This is why Aristotle posits that God’s sole activity is self-contemplation. Untangling a morass of false assertion, by contrast, is a lousy activity to absorb attention. Well-functioning trust relationships may also have non-instrumental value. Systems are better when they function properly, rather than deviantly. Treating assertions that *p* as evidence against *p* is dysfunctional.

Some disvalues are emergent. Dyads of distrust are bad; a prevalence is corrosive. Social institutions would dissolve; many would never have arisen. This likely includes the institution of testimony itself. Discussion allows people to develop skills, cultivate virtues, expand human understanding, perceive gaps in human knowledge, and develop appropriate humility. Learning together is more than adjusting confidence in isolated propositions. We help one another understand and interpret the world. Together we forge new conceptions and formulate better questions. We can be role models and inspire each other epistemically. By learning

together, communities bond. These values are threatened by widespread ranting, evidence-resistance, and a prevalence of distrust.

Morreau and Olsson conclude,

Excluding information resistant people from debates requires recognizing them as such, which realistically must consume resources. Censorship, de-platforming and other ways of excluding them furthermore violate social and democratic values of inclusiveness and free speech. Our study suggests that, in some contexts anyway, the epistemic benefits of excluding false ranters are not worth these costs. [...] It might be better simply to let everyone have their say.

But tracking false ranters, ignoring assertions, and treating testimony as counterevidence also have significant epistemic, social, moral, and opportunity costs. To apply simulations and formal measures of epistemic value to real life, these costs cannot be ignored.³

Acknowledgements

I am grateful to Catherine Elgin, Liz Camp, Hilary Kornblith, and Cat Saint-Croix for helpful comments, and to Jon Garthoff for discussion and feedback on an earlier draft. This research was supported by an ACLS Fellowship from the American Council of Learned Societies.

Notes

- 1 I am grateful to Catherine Elgin, Liz Camp, and Hilary Kornblith for insightful discussion and for suggesting the political examples.
- 2 See Georgi Gardiner (2022) "Attunement: On the Cognitive Virtues of Attention" *Social Virtue Epistemology*, eds. Alfano, Klein, and de Ridder. Routledge.
- 3 Morreau and Olsson write, "there are related studies suggesting [...] that restricting social information, or *equivalently*, [individuals'] being resistant to such information can sometimes be epistemically beneficial" (emphasis added). But restricting information (that is, censorship and similar) and an individual's not updating on that information are not equivalent. They constitute different socio-epistemic arrangements, require different epistemic institutions and cognitive resources, and generate different counterfactual epistemic conditions. If a simulation and measure of epistemic value cannot distinguish them, this redounds poorly on the model.

18c Commentary from Thi Nguyen

Michael Morreau and Erik J. Olsson’s “Learning from Ranters” makes a substantive social prescription based on computer modeling. Their argument proceeds through a demonstration, via a model, that sources of misinformation don’t actually affect open-minded epistemic agents. They conclude that *censorship of online sources of misinformation is often unwarranted*. But we must be quite sure that the presumptions of the model fit the facts of the world, if we are to accept that prescription.

Morreau and Olsson’s analysis focuses on “ranters” – epistemic agents who are closed-minded and don’t update their beliefs, but spout information into the world. Morreau and Olsson model two types of ranter – true ranters, who spout truth, and false ranters, who spout falsehoods. False ranters are, I take it, the model’s stand-in for real-world sources of misinformation. Morreau and Olsson then model the effect of false ranters on open-minded epistemic agents, in communicative networks. They conclude that false ranters don’t actually make things worse for open-minded epistemic agents. In fact, according to their model, the presence of false ranters can actually make open-minded agents epistemically better-off.

How could this be? In their models, the open-minded agents have other network connections to other informational sources. And, in their models, the majority of the sources in the network give true information. So open-minded agents can eventually figure out that the false ranters are false ranters. And because the false ranters are reliably false, open-minded agents can then use false ranters as something like lighthouses – they help you steer, by showing you what to avoid. If you can figure out that somebody is reliably wrong, then you can use them as a good guide to the truth. You just have to believe the opposite of whatever they say.

There seem to be three key presumptions in the model: first, that the relevant agents are open-minded. Second, the information and testimonial sources they have access to are true, in majority. Third, the false

ranters are reliably false. But, I worry these presumptions are inaccurate of the very situations of the world with which we should be most concerned.

In the model, open-minded agents continue to be open-minded when they hear the false ranter and continue to be in contact with the better, truth-conducive sources. That's how they figure out that the false ranter is false. From what I can tell, in the model, false ranters don't seem to be able to change the basic trust settings of the open-minded agent. But if we're talking about much of the really worrisome misinformation right now – Fox News, Breitbart, climate change denialists – that's not the right modeling presumption. What's actually going on, according to significant empirical research, is that the key sources of misinformation actually transform the trust settings of their audience, steering them away from all other sources of information. Kathleen Hall Jamieson and Frank Capella (2008) said that Rush Limbaugh and Fox News had set up an *echo chamber*, in which Rush Limbaugh had brought the member to systematically distrust all outsiders. And let's be clear – I mean “echo chamber” the original sense of the term: a social structure where members have been brought to *systematically distrust* all outside sources (Nguyen 2020). This dynamic has been confirmed by more recent research. According to Benkler et al.'s (2018) analysis of the 2016 American election, the right-wing media ecosystem created a *propaganda feedback loop*, in which audience members came to distrust mainstream news sources and only trust the constant confirmations they received from their preferred, right-wing media sources.

In other words, the really dangerous sources of misinformation in the real-world misinformation seem to have the capacity to *transform* open-minded people into closed-minded people and bring their audience members to distrust all those other, truth-conducive sources. Morreau and Olsson's model treats the audience as largely static – as essentially open-minded and in-contact with other good sources – in a way that diverges from how much actual misinformation seems to work.

Add to this one more feature. False ranters are easy to identify, and easy to usefully leverage, if they are reliably false. The model seems to rely on this fact. But this isn't how things work with actual real-world misinformation sources. Such sources provide a mixture of true claims and false claims. And, in many cases, strategically clever misinformation specialists will work to ensure that their true claims are relatively easy to verify, and their false claims relatively hard to.

To sum up: the model presumes that false ranters are constantly false, and that they don't alter the basic topology of trust, and the basic open-mindedness, of their audience. But both presumptions seem false of much real-world misinformation.

References

- Benkler, Y., Faris, R., & Roberts, H. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford: Oxford University Press.
- Jamieson, K.H. & Cappella, J.N. 2008. *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford: Oxford University Press.
- Nguyen, C.T. 2020. Cognitive Islands and Runaway Echo Chambers. *Synthese*, 197(7), 2803–2821.

T&F Proofs – Not for Distribution

18d Michael Morreau and Erik J. Olsson's Response to Commentaries

Our chapter in this volume reports preliminary findings from a study of the epistemic consequences of misinformation in social networks. We use the methodology of computer simulation, constructing computer models of Bayesian learning in networks that include untrustworthy sources. Our main result is that, under certain conditions, open-minded enquirers learn just as well in the presence of *false ranters*: information-resistant agents that repeatedly broadcast falsity within networks.

In real public fora, there are costs in excluding trolls, liars and bullshitters. Censorship and deplatforming take time and money, and they violate social and democratic norms of inclusiveness and free speech. We conjecture that, however urgent it might seem to shut down sources of misinformation, in many instances the epistemic benefits do not justify the costs.

Georgi Gardiner and C. Thi Nguyen in their discussions of our chapter call into question above all the realism of our model of social deliberation. We focus in this reply on what we take to be some of their main comments and criticisms.

Sometimes, we found, our simulated enquirers even *benefit* from the presence of false ranters, progressing more quickly toward full belief in the truth than they otherwise would. This learning involved *testimony inversion*. Enquirers came to recognize false ranters as such, and to treat their testimony as evidence to the contrary. Gardiner in her commentary recognizes that there are “cases where assertions can be a reason to believe the opposite” but claims that these are “marginal, require significant background evidence and context, and are about limited domains of assertion”. The implication is that testimony inversion can't be important in the real world, outside of our simulations.

Concerning the reality of testimony inversion, we note in the chapter the experiments reported by Collins, Hahn, von Gerber and Olsson (2018). They found that “[p]articipants used source reliability when assessing claim strength, and can consider sources to be anti-reliable or negatively correlated with the truth” (p. 9). Empirical evidence that testimony inversion is not only real but also very basic in human thinking comes from an earlier study by Lee and Cameron (2000) on

young children's ability to extract information from deceptive testimony. They observe (p. 16) that while "3-year-olds rejected the lie-teller's statement as reflecting his true state of knowledge and the true state of affairs", "[m]ost 4- and 5-year-olds had an even more advanced understanding":

They, with little or no prompting, correctly inferred the lie-teller's true knowledge state and the true state of affairs from the content of the lie-teller's lie. In other words, they understood that a lie, which is untruthful overall, may contain useful information about the lie-teller's belief as well as the true state of affairs.

Hence, rather than merely dismissing deceptive testimony as useless, even very young children can extract information from lies, including information about the true state of affairs. Evidently, testimony inversion is not a marginal phenomenon but is deeply rooted in human cognition.

Gardiner seems to think that the updating of trust in our model must place a heavy cognitive burden on open-minded enquirers:

Trust updating requires dynamic accountancy. One must track whose previous assertions conflicted with one's antecedent beliefs and weigh their future testimony accordingly. This effort is better invested elsewhere.

In fact, all our open-minded enquirers do is keep a log of how much they trust any given source at the time, updating as needed whenever a new message arrives from this source. The details are in our chapter. The cognitive burden presumably is quite manageable when the group of sources to be kept tabs on is small, say candidates in a political election. In online settings, where this group might be large, reputation-tracking software could in the future take on much of the work.

One of Nguyen's objections seems to rest on a misunderstanding. Referring to American politics, he argues that "the really dangerous sources of misinformation in the real-world misinformation seem to have the capacity to transform open-minded people into closed-minded people, and bring their audience members to distrust all those other, truth-conducive sources". Our model, by contrast, supposedly "treats the audience as largely static – as essentially open-minded and in-contact with other good sources – in a way that diverges from how much actual misinformation seems to work".

Our model does not in fact presume that false ranters "don't alter the basic topology of trust". In the course of a deliberation, which includes receiving information from false ranters, our open-minded agents almost always develop full belief. From then on, according to the laws of Bayesianism, they never change their minds. Just as Nguyen would like, then, in our model "sources of misinformation ... have the capacity to

transform open-minded people into closed-minded people". Indeed, of course, sources of *true* information also have this capacity! Misinformation that closes minds to reliable sources is not a focus of our present study, but it is a promising topic for further research using our model.

Another objection suggests a useful extension of our model. Real-life sources of misinformation, Nguyen argues, are harder to identify than our reliably false ranters. They "provide a mixture of true claims and false claims" and "strategically clever misinformation specialists will work to ensure that their true claims are relatively easy to verify, and their false claims relatively hard to". It is reasonable to expect, as Nguyen does, that keeping track of sources of misinformation will be harder when these sources strategically cover their tracks. He is furthermore right to think that studying this kind of deception is not possible in our current model, in which there is just a single claim under consideration. That would require an extension in which agents deliberate about several claims, some true and some false.

Several of Gardiner's and Nguyen's objections concern idealizations. We note in closing that scientific models need not be completely true to reality for them to generate accurate predictions. Idealizing away real but irrelevant features of target systems can furthermore improve explanations, by isolating factors that by themselves are sufficient to reproduce the phenomena of interest. Gardiner and Nguyen are right that some assumptions of our model are unrealistic. But whether these departures from reality undermine our results or strengthen them remains to be seen. Finding out requires empirical calibration of the model and testing its predictions in experiments with real social networks. We signal the need for such empirical work at the end of our chapter.

References

- Collins, P.J., Hahn, U., von Gerber, Y., and Olsson, E.J. (2018), "The Bi-directional Relationship between Source Characteristics and Message Content," *Frontiers in Psychology* 9(18). <https://doi.org/10.3389/fpsyg.2018.00018>
- Lee, K., and Cameron, C.A. (2000), "Extracting Truthful Information from Lies: Emergence of the Expression-Representation Distinction," *Merrill-Palmer Quarterly* 46(1): 1–20.

19 Education as the Social Cultivation of Intellectual Virtue

Michel Croce and Duncan Pritchard

1 Introductory Remarks

Educational theory is an obvious area of interest for social epistemology, not least because education clearly has some epistemic goals at its heart, and yet it is also naturally understood as an essentially social enterprise, one that all of us partake of in some form. Our interest in this chapter is in a specific conception of the epistemic goals of education, such that education is ultimately concerned, from an epistemic point of view at least, with the cultivation of intellectual character, and thus with the development of those intellectual character traits known as the intellectual virtues. In particular, how does thinking of the epistemic goals of education in this way inform our conception of education as an essentially social practice? As we will see, a key issue in this regard is the role that intellectual exemplars play within a virtue-theoretic account of the epistemology of education, and specifically the extent to which social interactions with these exemplars form part of this educational method.

2 Intellectual Character and the Epistemic Goals of Education

Education clearly serves many purposes, some of them social, some of them political, some of them practical, and so on. But a core aim of educational practice has to be epistemic. Indeed, it would be hard to understand how a set of practices could even count as educational unless they were geared towards epistemic goals. Of course, one could decree that henceforth one's educational system should be engaged in teaching nothing but falsehoods and propaganda. That would not be to adopt a revisionary style of education, however, but rather to give up on education altogether and pursue something different, in this case, indoctrination (even if one does so under the, now misleading, description of being 'education').

In any case, in what follows we will take it as given that a core aim of education is specifically epistemic.¹ That raises the further question of how these epistemic goals of education should be understood, which in

turn will have implications for how these educational practices should be conducted. Is the epistemic goal of education simply to get students to acquire true beliefs around a range of relevant subject matters (e.g., ones of societal utility)? If so, then learning answers by rote might make perfect sense as a pedagogical strategy, even if it results in individuals who may often fail to know what they (truly) believe (e.g., because they are never offered supporting reasons for their beliefs), much less understand it. Alternatively, perhaps the epistemic aim of education should be something more demanding, like knowledge (or at least justified true belief anyway) or understanding? Could such more elevated epistemic standings be acquired purely by rote learning? Possibly, though clearly this is far less obvious.

One influential way of thinking about the epistemic goals of education in the contemporary literature is not primarily in terms of the acquisition of an epistemic good, like knowledge, but rather in terms of the cultivation of *intellectual character*. On this conception, particular epistemic goods enter the picture in a secondary fashion, as being that which the development of intellectual character leads to.

One can see the attraction of putting the development of intellectual character into the heart of the educational enterprise. One wants students to be able to think for themselves, and that means an active engagement with the learning process, rather than merely coming to know lots of facts because the student has learnt them on good authority. Relatedly, having a developed intellectual character is a transferable skill, in that it means that once one has it one is better placed to be able to learn things for oneself, across a wide variety of new domains. Students who have acquired these skills will be in a position to acquire a range of epistemic goods like knowledge. Indeed, they will be particularly well placed to acquire elevated epistemic standings like understanding, the acquisition of which is usually held to depend on the active intellectual participation of the subject (i.e., as opposed to merely accepting the say-so of an authority).²

There are various ways that we can think about what the development of intellectual character in an educational context might mean. The standard way of conceiving of it in the literature is as the development of the *intellectual virtues*, where these are held to be the cognitive traits that collectively comprise intellectual character.³ The intellectual virtues are here understood along broadly Aristotelian lines, and hence are construed as more than simply a subject's reliable cognitive faculties and abilities.⁴ So this way of thinking about the educational development of intellectual character is not just a matter of giving students a certain kind of practical expertise, such as teaching them how to find things out, or teaching them certain practical skills, like critical thinking skills.⁵ More specifically, where these practical skills are taught, they are done

so in the service of developing specific intellectual virtues, rather than as ends in themselves.

Examples of the intellectual virtues include being intellectually humble, being observant, being intellectually conscientiousness, and being intellectually courageous. Consider being observant as an example. This is a more refined cognitive trait than simply having good perceptual skills. The latter may enable one to see clearly what is before one, and yet one might still fail to notice important features of the visual scene that only the observant person will detect. One can be born with good perceptual skills—such as one’s perceptual faculties—but the intellectual virtues are never innate. They must rather be cultivated, and indeed one needs to continue to cultivate them even once acquired, as otherwise they can be lost (so being intellectually virtuous is not like a skill such as riding a bike, where once learnt, it is rarely forgotten).

Another feature of the intellectual virtues that sets them apart from mere cognitive skills and faculties is that they involve a characteristic motivational state. This is, broadly understood, a desire for the truth, for getting things right. Cognitive skills need have no motivational state associated with them, and even if they do it needn’t be this kind of motivational state. For example, one could reliably manifest a cognitive skill for purely strategic reasons, but this is not possible for an intellectual virtue. Someone who acts as if they are intellectually humble in order to earn the plaudits of their peers, for example, is not actually manifesting this intellectual virtue at all.⁶

Intellectual virtues, like the virtues more generally, lie between two opposing vices, one of excess and one of deficiency (this is the ‘golden mean’). The challenge of acquiring an intellectual virtue involves having the good judgement to steer between these two vices in order to manifest the virtue. Consider, for example, the intellectual virtue of being intellectually courageous. The corresponding intellectual vice of deficiency would be intellectual cowardice, such as a failure to seek the truth because of the personal costs involved, like having to resist peer pressure. The corresponding intellectual vice of excess, in contrast, would be in manifesting the underlying cognitive traits to an immoderate degree. This would be a kind of intellectual rashness, where, for example, one takes undue intellectual risks, such as by ignoring opposing advice even when it is clearly relevant.

Navigating between the corresponding vices is particularly challenging given that it is usually accepted that there is no rubric that one can follow in order to manifest virtue, intellectual or otherwise. Instead, it is rather a matter of developing good judgement, which means, in turn, being sensitive to salient features of the situation and displaying

the appropriate motivational response to it, and that is acquired through observing, reflecting upon, and interacting with role models rather than studying a manual for virtuous behavior.⁷ In particular, although the standard account of virtue formation encompasses direct instruction into the vocabulary of the virtues as an initial stage, it then develops mainly through: (a) interaction with virtuous role models, and (b) the opportunity to identify and practice virtuous behavior in the specific settings that one finds oneself in.⁸

A final important aspect of the virtues, and thus the intellectual virtues, that is worthy of note is axiological. The idea is that they are constituent parts of a life of flourishing, and hence are intrinsically valuable. If that's right, then that would supply a further rationale for thinking that the epistemic goal of education should be the development of intellectual virtue, since this would be part of the wider goal of education to cultivate the virtues (i.e., both intellectual and non-intellectual), and thereby promote human flourishing.⁹

3 The Social Cultivation of Intellectual Virtue in Educational Settings

A feature of the intellectual virtues that is particularly relevant for our purposes is that their cultivation is essentially a social process, at least in terms of the acquisition of the intellectual virtue anyway (as opposed to its maintenance thereafter), which is the developmental stage that will be our focus here. This point is significant because while education is generally understood as an essentially social activity, it's not obvious why on other conceptions of the epistemic goal of education this should be the case, at least from a purely epistemic point of view at any rate. If the epistemic goal of education is merely to train students to have a certain set of true beliefs and cognitive skills, for example, then while as it happens the most efficient way of doing this at present is via social training, there is no obvious reason why future generations should be so limited. Perhaps there will be technological innovations that enable students to acquire these true beliefs and cognitive skills in isolation from others, or even be able to cognitively 'off-load' them to technology altogether?¹⁰ Of course, there might be other aspects of education besides its epistemic aspect that require social input, and which ensure that education is still an essentially social activity. But given the centrality of the epistemic goals of education to the educational enterprise, it would be at least surprising that from an epistemic point of view at least there is nothing essentially social about education.

Indeed, it isn't just that we tend to suppose that education is essentially social (and thus that the epistemic goals of education ought to be satisfied in a social manner too), but that we have a certain conception in mind of what the social component should involve. In particular, one

natural worry with the idea that epistemic goods like true belief might be the epistemic goal of education is that even if there is an essentially social dimension to educational practices, they might nonetheless be manifested in an entirely unidirectional fashion, such that the student simply defers to the educator. Instead, we take it that our natural conception of the social dimension of the educational enterprise involves active social engagement between both the educator and the student, such that the student simply cannot be a passive participant in this practice. As we will see, conceiving of the epistemic goals of education along virtue-theoretic lines speaks to both issues, in that the cultivation of the intellectual virtues is an essentially social process that requires an active contribution from the student.

As we've noted, the virtues are not innate, and so have to be acquired. Moreover, one cannot acquire them by oneself. One cannot acquire the virtues simply by reading a manual, for example. As previously indicated, there is no way of operationalizing the virtues into a set of rules that could be set out in a guidebook, as the manifestation of a virtue involves a highly refined sensitivity to the relevant features of the context, and this is not something that can be determined in advance of engaging within that context. What needs to be instilled into the subject are thus the right kinds of behavioral dispositions and the corresponding motivational states, and this is an essentially social process.

One's character is in general acquired through the manner in which one is embedded in social conditions, whereby children absorb behaviors and values from those around them, and in particular in response to their interactions with important adult figures in their lives, such as family members and teachers. At least some of these social interactions will be self-consciously understood by the adults as being directed at improving the child's character, and thus to this extent educational (though obviously this might not be explicitly characterized as such, and certainly there need be no mention of the intellectual virtues specifically). Think, for example, of how one guides a child's moral development, such as how a skillful teacher responds to conflict in the classroom to help those involved to see each other's point of view, or how she might deal with questions of fair play that arise in the playground. The goal is to cultivate certain kinds of dispositions and motivations, and that's to develop character in the broad sense that concerns us.¹¹

Our interest is specifically in the social development of the intellectual virtues, but one can see how this might arise within this kind of educational social setting (even if, as before, it is not self-consciously thought of this way by the agents involved). The idea we are exploring is normative rather than descriptive, however, in that not only is an intellectual character developed in such a scenario, but that a particular kind of intellectual character, one that comprises the intellectual virtues, ought to be developed in social educational contexts. Good educational

practice is accordingly reconceived in virtue-theoretic terms. Why is it important to educators that students are able to think for themselves rather than simply accepting claims on authority? Why do educators strive to stimulate an intrinsic desire for learning in their students, rather than simply making the case for the prudential value of education? And why do educators place such an emphasis on certain kinds of intellectual role models in their teaching? The thought is that we can make sense of these practices in terms of an implicit recognition that what education is trying to achieve is the cultivation of students' intellectual characters. So construed, education is a social practice that is, properly implemented, designed to cultivate the intellectual virtues.

One project of applied social epistemology that arises from this understanding of the epistemology of education is to consider ways in which explicitly characterizing educational activities in terms of the development of intellectual virtues might make those social practices more effective at achieving epistemic ends. So, for example, there have recently been projects that bring the intellectual virtues into schools, into prison education initiatives, and into University curricula.¹² Such projects present theoretical challenges, such as questions concerning the measurement of their efficacy, or whether the target should instead be the similar (though ultimately distinct) intellectual character traits associated with critical thinking.¹³ And of course, any appeal to intellectual virtues will encounter the general problems that face all virtue-theoretic proposals.¹⁴ Rather than work through these theoretical challenges here—which would require a wholesale defence of the centrality of the intellectual virtues to education—we want to instead explore a particular theoretical way of conceiving of how intellectual virtues are cultivated in educational settings that brings out its specifically social dimension.

4 Intellectual Exemplars

There has been a lot of work conducted on the role of exemplars in the development of virtuous character, including in educational settings, with the focus specifically on the moral virtues. The guiding idea behind *exemplarism* (as it is known) is that virtuous character is most naturally developed by emulating those that we admire, rather than trying to simply do the right thing in the abstract. Emulating the exemplar helps one to gain a better understanding of what appropriate conduct demands, and our attachment to the exemplar helps to motivate us to act as they would act.¹⁵ Rather than studying a manual for virtue—which as we saw above is simply unavailable—one instead acquires the virtues (at least in part) by observing virtuous role models in action and learning to mirror their virtuous behavior and motivations.¹⁶

While exemplarism has been widely explored with regard to the moral virtues, there hasn't been much discussion of how it would apply to the

intellectual virtues specifically, even though the same general principles should hold.¹⁷ The exemplars are meant to be a way of acquiring virtue in general, after all, rather than a particular kind of virtue. Accordingly, let's pursue the idea with the intellectual virtues expressly in mind. As we will see, the role played by intellectual exemplars in a virtue-theoretic account of education brings out one core way in which that account understands education as an essentially social activity.

Exemplars need not be perfect role models; indeed, it has been argued that perfect role-models—moral 'saints' for example—don't make good exemplars, precisely because they are so remote from ordinary folk, who are eminently fallible.¹⁸ Relatedly, intellectual exemplars needn't be intellectually virtuous in every respect; all that matters is that they exhibit certain kinds of intellectual virtues, which usually means at least one intellectual virtue to a high degree, and a cluster of related intellectual virtues to an above-average degree.¹⁹ So construed, students could be introduced to intellectual exemplars who are in some respects intellectually flawed, insofar as their intellectual character is admirable in relevant respects.

Somewhat surprisingly, an exemplar's imperfection can be beneficial to the whole process of developing virtuous habits in several ways. First, it makes it easier for students to associate specific role models with the particular character traits that make them intellectually exemplary, thereby providing students with a quick and manageable way to refer to—and distinguish among—specific intellectual virtues. Second, it provides the teacher with the opportunity to make the students work on character shortcomings by allowing them to reflect on the negative effects of an exemplar's intellectual flaws and their struggles to overcome those weaknesses. Finally, it counteracts the risk that the exceptionality of an exemplar's intellectual behavior discourages the students instead of motivating them to emulate it.

Exemplars can be introduced into the educational context *directly* or *indirectly*, depending on whether the students engage with the exemplar through direct social interaction or only indirectly by learning about them. A key advantage of direct interaction with intellectual exemplars is that the students can see intellectual virtues exercised in a context with which they are already familiar. Thus, this form of interaction speeds up the students' assessment of the benefits of virtuous behavior and offers them a concrete trajectory to replicate the exemplar's behavior—two features that it is much harder to secure through indirect interaction with an intellectual exemplar.

Typical cases of intellectual exemplars with whom students have direct social interaction include other students and the teacher. The intellectual exemplarity of one's peers in educational settings is the paradigmatic form of imperfect exemplarity. A student might display a good deal of intellectual courage by speaking up on behalf of the last

to speak to let everyone in the classroom know that they should value more his or her opinion. Such virtuous behavior is compatible with further less-than-virtuous features of the student's intellectual profile, but it nonetheless provides the other classmates with a luminous example of how exercising the virtues can have an impact on the social environment they live in.

One might suspect that in the early stages of the development of an intellectual character students will lack the ability to evaluate a case like this on their own: they might surely be impressed by their courageous classmate, but the steps from admiring him or her to recognizing the virtuousness of such behavior and desiring to emulate it require the teacher's support. Far from constituting an obstacle to the argument, this sheds light on the function the teacher performs in terms of guiding the intellectual development of the students. This task already requires that the teacher be somewhat sensitive to the students' epistemic needs, intellectually empathic, and practically wise (among other things): thus, the more intellectually virtuous a teacher is the more likely it is that she can help students build an intellectual character by developing the appropriate emotional reactions and intellectual motivations towards instances of exemplary behavior.

Further educational advantages arise from direct social interaction between the students and an exemplary teacher, that is, a teacher who exemplifies some virtue in her activity in the class (a possibility that is considerably more feasible if one rejects a 'saintly' conception of exemplars). As we have just seen, it is not a necessary requirement of exemplars that they perform a guiding role—one's classmates can manifest virtue but surely cannot guide other fellows in developing an intellectual character. So, where an exemplar like a teacher is playing this guiding role, then this reinforces the educational function that the exemplar is playing. The student is not merely seeing how the exemplar behaves in relevant conditions—their (in all likelihood partial) manifestation of intellectual virtues—but is also being explicitly guided by the exemplar in her own intellectual development. Moreover, the direct involvement with the exemplar increases the scope for emotional and intellectual 'contagion' (as it is known), whereby the student, by closely identifying with and interacting with the exemplar, is able to transform their own intellectual responses and motivations and thereby come closer to developing intellectual virtue herself.²⁰

Exemplars can also be introduced into the educational setting indirectly. This is when students are asked to study and reflect on figures who have manifested intellectually virtuous character traits. For example, students might be tasked to study an important historical figure who has an impressive intellectual pedigree, such as a pioneering scientist, or a reforming politician. Interestingly, these indirect exemplars needn't be actual, as fictional figures can also play this role. By immersing oneself

in, say, a novel, and studying it closely, a student might gain a deep appreciation of one of the characters involved, and thereby gain insight into the nature of intellectual virtue.

One clear advantage of indirect use of exemplars through narratives is that the narrator has the possibility to provide a detailed description of the path that the exemplar has followed to become intellectually virtuous. This might encompass a description of the struggles the exemplar had to go through, the obstacles she had to overcome, and the personal and social benefits of her exemplary behavior. Moreover, narratives are somewhat stable and therefore allow one to engage with indirect exemplarity over time, to see whether one still admires an exemplar's intellectual behavior and how close one got to the exemplar after attempting to emulate her intellectual deeds. Both these features are hard to find in direct exemplars, in that one typically encounters them during a specific period and might lack the resources or the opportunity to figure out how the exemplars got to be who they are.

Both direct and indirect use of exemplars is found in educational initiatives that are focused on the intellectual virtues. A contemporary project that has brought the intellectual virtues into the heart of the curriculum of two schools in the United States, for example, involves training educators in the nature of the intellectual virtues so that they can act as exemplars for their students (direct exemplars) and also highlights intellectual role models from history and literature (indirect exemplars).²¹ Similarly, a current educational initiative devoted to bringing the intellectual virtues into a US University curriculum involves, *inter alia*, highlighting the intellectual virtues by focusing both on important historical figures and fictional characters who have manifested particular intellectual virtues (indirect exemplars) and showcasing profiles of local faculty who students are able to directly interact with (direct exemplars).²²

Since exemplars can be both indirect and direct, and only the latter requires actual social interaction between the student and the educator, then one might think that the social dimension to implementing a virtue-theoretic approach to education is optional. In particular, why can't one develop one's intellectual character entirely in isolation by simply engaging with indirect exemplars, thereby doing without the need for any social interaction?

One issue here is that even when it comes to indirect exemplars there is a need for students to be guided by the skilled educator in terms of how to respond to the exemplar. Indeed, this is especially the case when it comes to indirect exemplars given the lack of social interaction between the student and the exemplar. Remember that our focus is on students who are in the process of developing an intellectual character. How else is such a student to learn anything from an indirect exemplar except by being guided in their engagement with the exemplar? There is thus still a need for social interaction as part of the educational methodology.

The more important point, however, is that while indirect exemplars have a role to play in the acquisition of intellectual character on the virtue-theoretic proposal, they are no substitute for direct exemplars. As a number of commentators have noted, the effectiveness of exemplars in developing virtue (intellectual or otherwise), especially when it comes to the acquisition of virtue (i.e., as opposed to a later stage where the virtue is merely being cultivated), depends upon a range of factors. These factors are overwhelmingly present, however, or at least present to a higher degree, in direct as opposed to indirect exemplars.

We have already noted that it is generally accepted that exemplars should not be paragons of virtue because that makes them too distant from people who are meant to learn from them. The general principle in play here is that exemplars need to be people that the student can identify with, which means that while they must be clearly superior to the student along some relevant axes of evaluation, they cannot be so dissimilar that the student simply finds them (and their behavior, values, and so on) alien. For example, Michel Croce and Maria Silvia Vacca-rezza (2017) defend the importance of close-by ordinary exemplars. In particular, they argue that moral heroes (who might well be morally flawed in various respects), as opposed to moral saints (who lack such moral flaws), can be more effective as exemplars because their very accessibility aids imitability.

In a similar fashion, Meira Levinson (2012, ch. 4) has emphasized the idea of ‘life-sized’, rather than ‘out-sized’, role models, where this means not just exemplars who are more like heroes than saints, but also exemplars who are also similar to the students in other respects—she lists “ethnicity or race, culture, religion, national origin, residence, or class” (160) as relevant considerations in this regard. Levinson argues that such life-sized role models who the students can relate to are better able to inspire the development of virtue than more conventional role models (where she has the civic virtues specifically in mind), especially since the latter are often not similar to the students in the relevant respects (as conventional role models are more likely to be male, white, and from a more privileged class background).²³

This line of reasoning is not just intuitively plausible, but has also been supported by some recent empirical work. This suggests that when it comes to moral education exemplars who are relatable to the student group are much more effective at generating relevant changes in moral behavior than exemplars who are judged to be very different from that group, such as distant historical figures.²⁴ Although this study is focused on moral exemplars, one would anticipate similar results in the case of intellectual exemplars.

Of course, one might counter that the foregoing merely indicates that the exemplars should be viewed as someone that the students can identify with, which doesn’t preclude the possibility that the exemplars are

nonetheless indirect. Perhaps, for example, one should simply replace the use of historical or fictional exemplars that students struggle to relate to with exemplars that are more accessible (and so more diverse in terms of ethnicity, class, gender, and so on)? Notice, however, that if the relatability of the exemplars is so important in this regard, then one would naturally expect this to entail that direct exemplars will be in general more effective than indirect exemplars, for the simple reason that the social interaction between the exemplar and the students is likely to make that exemplar more relatable to the students.

This last point relates to a further important issue in this regard, which is the extent to which direct exemplars, precisely because of their proximity to the students, and their regular social interaction with them, are much better placed to aid students in their development of intellectual virtue.²⁵ Proximity and interaction are clearly going to be tremendously helpful when it comes to reinforcing the kind of habitual change that is crucial to virtue development. In particular, it will generate positive feedback loops of encouragement when the behavior and/or motivations are apt and discouragement when they are not, along with the possibility of the kind of emotional contagion noted above. The interactions with the student will, after all, be individualized to them, since they involve a direct engagement that is lacking when it comes to using indirect exemplars. This allows for a kind of bespoke learning environment, with experiences and projects that are shared by both student and exemplar, thereby reinforcing the positive feedback loops of virtue development. Indeed, some commentators have gone so far as to emphasize the importance of a kind of friendship between the student and the educator in this regard.²⁶

A further consideration in this regard is that the role of direct exemplars is arguably more important when it comes to intellectual exemplars than moral exemplars. This is because the latter is a much more familiar category. Consider the virtue of being morally courageous, for example, as compared to the corresponding virtue of being intellectually courageous. Few would struggle to come up with examples of people who have instantiated the former, but many would surely find it much more difficult to list people who have instantiated the latter. This point isn't restricted to this intellectual virtue either, as it seems that most intellectual virtues are harder to recognize than their moral counterparts. It wouldn't be difficult to give examples of people who have the virtue of humility, for example, but giving corresponding examples of people who have the virtue of intellectual humility would be much harder.

The crux of the matter is that our practices are already shot-through with moral talk and instances of morally praiseworthy behavior—and also instances of morally lacking behavior too, of course. This makes it easier for us to be able to use indirect exemplars to guide someone's development of moral virtue, as there is a common background of moral

examples to attach one's use of the exemplars to. Since the intellectual virtues are not already represented in our practices to the same extent, however, then that makes appealing to indirect exemplars much more difficult. In particular, it will be much easier to guide someone's intellectual development by employing direct exemplars; for example, by actually putting the student into contact with someone who is intellectually humble, and exploring what this means in practice.

5 Concluding Remarks

We've explored the idea of educational theory as applied social epistemology by considering the prominent proposal that the epistemic goal of education is the development and cultivation of intellectual character, and thus the intellectual virtues that constitute one's intellectual character. As we've seen, one way of bringing out the essentially social nature of education, so conceived, is by considering the importance of intellectual exemplars to such an educational strategy. In particular, while we've noted that there can be ways of employing intellectual exemplars that needn't involve social interaction, the most potent use of intellectual exemplars in this regard will be as embedded within social interactions with the student.

Notes

- 1 For an overview of the contemporary literature on the epistemology of education, see Robertson (2009) and Baehr (2016).
- 2 For further discussion of understanding on this front, and in particular how understanding can be more demanding to acquire than the corresponding knowledge, see Kvanvig (2003), Grimm (2006), Pritchard (2009, 2014) and Pritchard, Millar and Haddock (2010, ch. 4), and Greco (2013). For a defence of the axiological importance of first-hand knowledge and understanding, see Pritchard (2016).
- 3 See Hyslop-Margison (2003), Battaly (2006), MacAllister (2012), Sockett (2012), Pritchard (2013, 2015, 2018, 2020, forthcoming), Byerly (2019), and the essays collected in Baehr (2015). For some contemporary treatments of the intellectual virtues, see Zagzebski (1996), Roberts and Wood (2007), Pritchard, Millar and Haddock (2010), and Baehr (2011). For an overview of the contemporary literature on this topic, see Battaly (2014) and Turri, Alfano and Greco (2017).
- 4 Note that there is a contemporary proposal that thinks of the intellectual virtues in ways that encompasses a subject's reliable cognitive faculties and cognitive skills (and which thus departs from the Aristotelian model), but such virtue reliabilism, as it is sometimes known, is not the view that concerns us here. For discussion of such a proposal, see Sosa (1991) and Greco (1999).
- 5 There is a lively debate in the literature about whether an intellectual character-based education should be aiming at the development of the intellectual virtues or merely at certain critical thinking capacities. See, for example, Siegel (1988, 1997, 2017, 2017), Hyslop-Margison (2003), Huber

- and Kuncel (2016), Hitchcock (2018), Baehr (2019), Carter, Kotzee and Siegel (2019), and Pritchard (forthcoming).
- 6 The differences between the intellectual virtues and cognitive skills more generally have led some commentators to argue that the former are not plausible candidates for extended cognition, in contrast to the latter. This has a bearing on the epistemology of education, given the important social scaffolding involved in educational practices, and also the increasingly prominent dependence on technology in teaching. For discussion of these issues, see Pritchard (2015, 2018). See also endnote 10.
 - 7 Virtue theory as it is normally understood thus goes hand-in-hand with a kind of particularism about good conduct. (Note that particularism is usually cast in terms of good moral conduct specifically, but here we are using it as it applies to good conduct generally). For a classic discussion of this point, see McDowell (1979).
 - 8 For a discussion of the standard approach to virtue formation, see Porter (2016).
 - 9 For developments of this kind of line with regard to the role of the virtues in education, see Carr (2014) and Kristjánsson (2015).
 - 10 One issue that is relevant here is the extent to which technology that is employed in education—or even social features of the educational context—might become, over time, an extended cognitive process on the part of the student, in the sense famously articulated by Clark and Chalmers (1998). For further discussion of the implications of extended cognition for educational practice, see Pritchard (2015, 2018), Carter and Pritchard (2017), Heersmink and Knight (2018), Kotzee (2018), and English, Ravenscroft and Pritchard (2021). For further discussion of the more general epistemological issues raised by cognitive augmentation, see Carter and Pritchard (2019).
 - 11 There is a wide-ranging literature on the relationship between education and the development of the moral virtues specifically, though this is at least partly orthogonal to our present concerns, which are specifically regarding the epistemic goals of education (and thus the intellectual virtues). See, for example, Carr (2014) and Kristjánsson (2015).
 - 12 Baehr's (2015) work in developing school curricula based around the intellectual virtues has been very influential in this respect. See Pritchard (2019, 2021) for discussion of a prison education initiative focused on developing the intellectual virtues, and see Orona and Pritchard (2021) for discussion of a pilot project bringing the intellectual virtues into the heart of a university curriculum. See also the *Self, Virtue and Public Life* project, led by Nancy Snow at the University of Oklahoma, which is devoted to bringing the civic virtues into the university curriculum (<https://selfvirtueandpubliclife.com>).
 - 13 On the issue of the difficulty of measuring the effectiveness of educational interventions involving the intellectual virtues, see Curren and Kotzee (2014), Kotzee (2015), and Carter, Kotzee and Siegel (2019). For discussion of the relative merits of educational strategies that focus on the intellectual virtues or critical thinking skills, see the literature listed in endnote 5. For some of the empirical literature regarding the effectiveness of critical thinking-based educational strategies, see Arum and Roska (2010), Seifert, Goodman, King and Baxter Magolda (2010), Liu, Mao, Frankel and Xu (2016), Liu, Liu, Roohr and McCaffrey (2016), and Roohr, Liu and Liu, (2016). For some of the empirical literature that is more relevant to intellectual virtue-based educational strategies, see Litman and Spielberger (2003), Krumrei-Mancuso and Rouse (2016), Lins de Holanda Coelho, Hanel and Wolf (2018), and Orona and Pritchard (2021).

- 14 To take one prominent example, there is the situationist critique of virtue theory found in, for example, Harman (1999, 2000) and Doris (2002). See also the application of this situationist critique to the intellectual virtues specifically found in Alfano (2012). For some responses to the latter, including with the epistemology of education specifically in mind, see Pritchard (2015), Baehr (2017), and Carter and Pritchard (2017).
- 15 For an influential recent discussion of exemplarism and its role in the development of virtue, see Zagzebski (2017). See also Zagzebski (2010). For a recent discussion of moral exemplars in the Confucian (as opposed to Aristotelian) tradition, see Olberding (2012). For discussion of exemplarism specifically in the educational context, see Porter (2016), Croce and Vaccarezza (2017), Croce (2019, 2020b), and Korsgaard (2019). For a recent expression of skepticism about exemplarism as an educational strategy, albeit focused on particular virtues and concerned with students who have already developed a (viceful) cognitive character, see Tanesini (2016). See also Alfano and Sullivan (2019), which questions whether standard forms of exemplarism can be employed to combat testimonial injustice.
- 16 Croce (2019, 2020a) unpacks the educational stages involved in exemplarism into what he refers to as the *exemplarist dynamic*, where the three stages are natural admiration, conscientious reflection, and emulation.
- 17 For an exception, see Croce (2020a, ch. 7). See also Alfano and Sullivan's (2019) discussion of negative epistemic exemplars, which focuses on the potential role of exemplars with regard to combating epistemic injustice, and Tanesini (2016), which examines how exemplars might be problematic with regard to some specific educational projects. In general, where the idea of epistemic exemplars does get discussed, it tends to be in passing, as part of a wider discussion of exemplars. See, for example, Baehr's (2011, ch. 8) remarks on the narrative of the crystallographer, Medina's (2013, ch. 5) discussion of epistemic heroes, van Dongen's (2017) comments on Albert Einstein, and Zagzebski's (2017, *passim*) discussion of the sage.
- 18 This point is usually made in the moral domain regarding moral saints—see, especially, Wolff (1982)—but the point is equally applicable in the intellectual domain. For a recent defence of a liberal approach to exemplars that allows a broad range of cognitive subjects to play this role, see Croce (2020b). See also Baehr (2015, ch. 13), who argues for a 'realistic' employment of exemplars in the classroom.
- 19 This assumes, of course, that one is rejecting the unity of the virtues thesis, usually attributed to Aristotle, that in order to have one virtue one must have them all. For a helpful critical discussion of this thesis, see Wolff (2007).
- 20 For more on the notion of emotional contagion and its role in virtue development, see Kristjánsson (2015, 2018, 2020). For an important empirical study of emotion contagion, as a three-stage process involving mimicry, feedback, and contagion, see Hatfield, Cacioppo and Rapson (1993).
- 21 This is the *Educating for Intellectual Virtues* project that was run by Jason Baehr (<https://intellectualvirtues.org>). This primarily led to bringing the intellectual virtues into the curriculum of the recently-founded Intellectual Virtues Academy of Long Beach, a charter middle school in California, but has also influenced the Intellectual Virtues Academy high school, also in Long Beach. See Baehr (2015) for an overview of the intellectual basis for the project.
- 22 This is the *Anteater Virtues* project led by one of the present authors at the University of California, Irvine. The project is described in detail in Orona and Pritchard (2021).

- 23 Making use of ‘life-sized’ role models in this way might respond to some of the worries about the educational employment of exemplars raised by Tanesini (2016), which in part concern the fact that students might not be inclined to appropriately respond to the exemplar. As we have noted, this might very much depend on how relatable the exemplars chosen are. Interestingly, the intellectual exemplars chosen as part of the *Anteater Virtues* project at the University of California, Irvine were selected with diversity in mind (especially racial and gender diversity) for just this reason. The empirical study of this initiative, described in Orona and Pritchard (2020), shows that the developmental improvement in the students who participated in this project was consistent across all the main student demographics, including female and URM (under-represented minorities).
- 24 See Han, Kim, Jeong and Cohen (2017). See also the empirical work noted in endnote 23.
- 25 This point is also emphasized by Levinson (2012, 160), who argues that exemplars work best when part of an “active, relationship-orientated, and experiential approach”.
- 26 See, for example, Kristjánsson (2020).

References

- Alfano, M. (2012). ‘Extending the Situationist Challenge to Responsibility Virtue Epistemology’, *Philosophical Quarterly* 62, 223–249.
- Alfano, M., & Sullivan, E. (2019). ‘Negative Epistemic Exemplars’, *Overcoming Epistemic Injustice: Social and Psychological Perspectives*, (eds.) S. Goguen & B. Sherman, 17–32, Lanham, MD: Rowman & Littlefield.
- Arum, R., & Roska, J. (2010). *Academically Adrift: Limited Learning on College Campuses*, Chicago, IL: Chicago University Press.
- Baehr, J. (2011). *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*, Oxford: Oxford University Press.
- (2015). *Cultivating Good Minds: A Philosophical & Practical Guide to Educating for Intellectual Virtues*, (available at: <https://intellectualvirtues.org/why-should-we-educate-for-intellectual-virtues-2-2/>).
- (ed.) (2016). *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*, London: Routledge.
- (2017). ‘The Situationist Challenge to Educating for Intellectual Virtues’, *Epistemic Situationism*, (eds.) M. Alfano & A. Fairweather, 192–215, Oxford: Oxford University Press.
- (2019). ‘Intellectual Virtues, Critical Thinking, and the Aims of Education’, *Routledge Handbook of Social Epistemology*, (eds.) P. Graham, M. Fricker, D. Henderson, N. Pedersen & J. Wyatt, 447–457, London: Routledge.
- Battaly, H. (2006). ‘Teaching Intellectual Virtues: Applying Virtue Epistemology in the Classroom’, *Teaching Philosophy* 29, 191–222.
- (2014). ‘Intellectual Virtues’, *Handbook of Virtue Ethics*, (ed.) S. van Hoof, 177–187, London: Acumen.
- Byerly, T. R. (2019). ‘Teaching for Intellectual Virtue in Logic and Critical Thinking Classes: Why and How’, *Teaching Philosophy*, (Online First: <https://doi.org/10.5840/teachphil201911599>).
- Carr, D. (2014). *Educating the Virtues: An Essay on the Philosophical Psychology of Moral Development and Education*, London: Routledge.

- Carter, J. A., Kotzee, B., & Siegel, H. (2019). 'Educating for Intellectual Virtue: A Critique from Action Guidance', *Episteme*, (Online First: <https://doi.org/10.1017/epi.2019.10>).
- Carter, J. A., & Pritchard, D. H. (2017). 'Epistemic Situationism, Epistemic Dependence, and the Epistemology of Education', *Epistemic Situationism*, (eds.) M. Alfano & A. Fairweather, 168–191, Oxford: Oxford University Press.
- (2019). 'The Epistemology of Cognitive Enhancement', *Journal of Medicine & Philosophy* 44, 220–242.
- Clark, A., & Chalmers, D. (1998). 'The Extended Mind', *Analysis* 58(1), 7–19.
- Croce, M. (2019). 'Exemplarism in Moral Education: Problems with Applicability and Indoctrination', *Journal of Moral Education* 48, 291–302.
- (2020a). *Epistemic Inequality Reconsidered: An Inquiry into Epistemic Authority*, PhD Dissertation, University of Edinburgh.
- (2020b). 'Moral Exemplars in Education: A Liberal Account', *Ethics & Education* 15, 186–199.
- Croce, M., & Vaccarezza, M. S. (2017). 'Educating through Exemplars: Alternative Paths to Virtue', *Theory and Research in Education* 15, 5–19.
- Curren, R., & Kotzee, B. (2014). 'Can Virtue Be Measured?', *Theory and Research in Education* 12, 266–282.
- Doris, J. (2002). *Lack of Character: Personality and Moral Behavior*, Cambridge: Cambridge University Press.
- English, A., Ravenscroft, J., & Pritchard, D. H. (2021). 'Assistive Technology and Extended Cognition', *Synthese*, (Online First, <https://doi.org/10.1007/s11229-021-03166-9>).
- Greco, J. (1999). 'Agent Reliabilism', *Philosophical Perspectives* 13, 273–296.
- (2013). 'Episteme: Knowledge and Understanding', *Virtues and Their Vices*, (eds.) K. Timpe & C. Boyd, ch. 13, Oxford: Oxford University Press.
- Grimm, S. (2006). 'Is Understanding a Species of Knowledge?', *British Journal for the Philosophy of Science* 57, 515–535.
- Han, H., Kim, J., Jeong, C., & Cohen, G. (2017). 'Attainable and Relevant Moral Exemplars are More Effective than Extraordinary Exemplars in Promoting Voluntary Service Engagement', *Frontiers in Psychology* 8, 283.
- Harman, G. (1999). 'Moral Philosophy Meets Social Psychology: Virtue Ethics and the Fundamental Attribution Error', *Proceedings of the Aristotelian Society* 119, 316–331.
- (2000). 'The Nonexistence of Character Traits', *Proceedings of the Aristotelian Society* 100, 223–226.
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). 'Emotional Contagion', *Current Directions in Psychological Science* 2, 96–99.
- Heersmink, R., & Knight, S. (2018). 'Distributed Learning: Educating and Assessing Extended Minds', *Philosophical Psychology* 31, 969–990.
- Hitchcock, D. (2018). 'Critical Thinking', *Stanford Encyclopedia of Philosophy*, (ed.) E. N. Zalta, <https://plato.stanford.edu/entries/critical-thinking/>.
- Huber, C. R., & Kuncel, N. R. (2016). 'Does College Teach Critical Thinking? A Meta-Analysis', *Review of Educational Research* 86, 431–468.
- Hyslop-Margison, E. (2003). 'The Failure of Critical Thinking: Considering Virtue Epistemology as a Pedagogical Alternative', *Philosophy of Education Society Yearbook* 2003, 319–326.

- Korsgaard, M. T. (2019). 'Exploring the Role of Exemplarity in Education: Two Dimensions of the Teacher's Task', *Ethics & Education* 14, 271–284.
- Kotzee, B. (2015). 'Problems of Assessment in Educating for Intellectual Virtue', *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*, (ed.) J. Baehr, 142–160, London: Routledge.
- (2018). 'Cyborgs, Knowledge and Credit for Learning', *Extended Epistemology*, (eds.) J. A. Carter, A. Clark, J. Kallestrup, S. O. Palermos & D. H. Pritchard, 221–238, Oxford: Oxford University Press.
- Kristjánsson, K. (2015). *Aristotelian Character Education*, London: Routledge.
- (2018). *Virtuous Emotions*, Oxford: Oxford University Press.
- (2020). 'Aristotelian Character Friendship as a 'Method' of Moral Education', *Studies in Philosophy and Education* 39, 349–364.
- Krumrei-Mancuso, E. J., & Rouse, S. V. (2016). 'The Development and Validation of the Comprehensive Intellectual Humility Scale', *Journal of Personality Assessment* 98, 209–221.
- Kvanvig, J. L. (2003). *The Value of Knowledge and the Pursuit of Understanding*, Cambridge: Cambridge University Press.
- Levinson, M. (2012). *No Citizen Left Behind*, Cambridge, MA: Harvard University Press.
- Lins de Holanda Coelho, G., Hanel, P. H. P., & Wolf, L. J. (2018). 'The Very Efficient Assessment of Need for Cognition: Developing a Six-Item Version', *Assessment*, <https://doi.org/10.1177/1073191118793208>.
- Litman, J. A., & Spielberger, C. D. (2003). 'Measuring Epistemic Curiosity and Its Diverse and Specific Components', *Journal of Personality Assessment* 80, 75–86.
- Liu, O. L., Liu, H., Roohr, K. C., & McCaffrey, D. F. (2016). 'Investigating College Learning Gain: Exploring a Propensity Score Weighting Approach', *Journal of Educational Measurement* 53, 352–367.
- Liu, O. L., Mao, L., Frankel, L., & Xu, J. (2016). 'Assessing Critical Thinking in Higher Education: The HEIghten™ Approach and Preliminary Validity Evidence', *Assessment & Evaluation in Higher Education* 41, 677–694.
- MacAllister, J. (2012). 'Virtue Epistemology and the Philosophy of Education', *Journal of Philosophy of Education* 46, 251–270.
- McDowell, J. (1979). 'Virtue and Reason', *The Monist* 62, 331–350.
- Medina, J. (2013). *The Epistemology of Resistance. Gender and Racial Oppression, Epistemic Injustice, and Resistant Imagination*, Oxford: Oxford University Press.
- Olberding, A. (2012). *Moral Exemplars in the Analects: The Good Person Is That*, London: Routledge.
- Orona, G. A., & Pritchard, D. H. (2021). 'Inculcating Curiosity: Pilot Results of an Online Module to Enhance Undergraduate Intellectual Virtue', *Assessment & Evaluation in Higher Education*, (Online First, <https://doi.org/10.1080/02602938.2021.1919988>).
- Porter, S. L. (2016). 'A Therapeutic Approach to Intellectual Virtue Formation in the Classroom', *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*, (ed.) J. Baehr, 221–239, London: Routledge.
- Pritchard, D. H. (2009). 'Knowledge, Understanding and Epistemic Value', *Epistemology (Royal Institute of Philosophy Lectures)*, (ed.) A. O'Hear, 19–43, Cambridge: Cambridge University Press.

- (2013). 'Epistemic Virtue and the Epistemology of Education', *Journal of Philosophy of Education* 47, 236–247.
- (2014). 'Knowledge and Understanding', *Virtue Scientia: Bridges between Virtue Epistemology and Philosophy of Science*, (ed.) A. Fairweather, 315–328, Dordrecht, Holland: Springer.
- (2015). 'Intellectual Virtue, Extended Cognition, and the Epistemology of Education', *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology*, (ed.) J. Baehr, 113–127, London: Routledge.
- (2016). 'Seeing It for Oneself: Perceptual Knowledge, Understanding, and Intellectual Autonomy', *Episteme* 13, 29–42.
- (2018). 'Neuromedia and the Epistemology of Education', *Metaphilosophy* 49, 328–349.
- (2019). 'Philosophy in Prisons: Intellectual Virtue and the Community of Philosophical Inquiry', *Teaching Philosophy*, (Online First: <https://doi.org/10.5840/teachphil201985108>).
- (2020). 'Educating for Intellectual Humility and Conviction', *Journal of Philosophy of Education* 54, 398–409.
- (2021). 'Philosophy in Prison and the Cultivation of Intellectual Character', *Journal of Prison Education and Reentry* 7(2), 130–143.
- (Forthcoming). 'Cultivating Intellectual Virtues', *Routledge Handbook of Philosophy of Education*, (ed.) R. Curren, London: Routledge.
- Pritchard, D. H., Millar, A., & Haddock, A. (2010). *The Nature and Value of Knowledge: Three Investigations*, Oxford: Oxford University Press.
- Roberts, R. C., & Wood, W. J. (2007). *Intellectual Virtues: An Essay in Regulative Epistemology*, Oxford: Oxford University Press.
- Robertson, E. (2009). 'The Epistemic Aims of Education', *Oxford Handbook of Philosophy of Education*, (ed.) H. Siegel, 11–34, Oxford: Oxford University Press.
- Roohr, K. C., Liu, H., & Liu, O. L. (2016). 'Investigating Student Learning Gains in College: A Longitudinal Study', *Studies in Higher Education* 42, 2284–2300.
- Seifert, T. A., Goodman, K., King, P. M., & Baxter Magolda, M. B. (2010). 'Using Mixed Methods to Study First-Year College Impact on Liberal Arts Learning Outcomes', *Journal of Mixed Methods Research* 4, 248–267.
- Siegel, H. (1988). *Educating Reason: Rationality, Critical Thinking, and Education*, New York: Routledge.
- (1997). *Rationality Redeemed? Further Dialogues on an Educational Ideal*, New York: Routledge.
- (2017). *Education's Epistemology: Rationality, Diversity, and Critical Thinking*, Oxford: Oxford University Press.
- Sockett, H. (2012). *Knowledge and Virtue in Teaching and Learning: The Primacy of Dispositions*, London: Routledge.
- Sosa, E. (1991). *Knowledge in Perspective: Selected Essays in Epistemology*, Cambridge: Cambridge University Press.
- Tanesini, A. (2016). 'Teaching Virtue: Changing Attitudes', *Logos & Episteme* 7, 503–527.
- Turri, J., Alfano, M., & Greco, J. (2017). 'Virtue Epistemology', *Stanford Encyclopedia of Philosophy*, (ed.) E. N. Zalta, <https://plato.stanford.edu/entries/epistemology-virtue/>.

- van Dongen, J. (2017). 'The Epistemic Virtues of the Virtuous Theorist: On Albert Einstein and His Autobiography', *Epistemic Virtues in the Sciences and the Humanities*, (eds.) J. van Dongen & H. Paul, 63–77, Dordrecht, Holland: Springer.
- Wolf, S. (1982). 'Moral Saints', *Journal of Philosophy* 79, 419–439.
- (2007). 'Moral Psychology and the Unity of the Virtues', *Ratio* 20, 145–167.
- Zagzebski, L. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, Cambridge: Cambridge University Press.
- (2010). 'Exemplarist Virtue Theory', *Metaphilosophy* 41, 41–57.
- (2017). *Exemplarist Moral Theory*, Oxford: Oxford University Press.

T&F Proofs – Not for Distribution

19b Commentary from Alessandra Tanesini

Intellectual Virtue, Imitation, and Education

Education, in formal and informal settings, is a significant feature of human communities. As Croce and Pritchard acknowledge educative practices and organizations serve many purposes: civic, moral, and epistemic. In their contribution to this volume, Croce and Pritchard focus on the epistemic goal of education which they identify as the cultivation of the intellectual character of the students via the acquisition of intellectual virtues. These virtues are constitutive of intellectual flourishing and include intellectual humility, courage, open-mindedness, inquisitiveness, independence of thought, and the disposition of being observant. Croce and Pritchard also single out one educational strategy in the pursuit of this epistemic goal: direct and indirect exposure to real or fictional less-than-perfect intellectual exemplars under the guidance of teachers.

Croce and Pritchard thus present education as a set of social epistemic practices aiming at an individualistic goal. This aim is the acquisition by each student of those epistemic character traits that contribute to his or her intellectual flourishing. In my view, this account is insufficiently social. It also characterizes contemporary educational practices in a manner that pays insufficient attention to their continuities with the epistemic goals of social or cultural learning as practiced in human societies prior to the advent of state-funded education and modernization.

Current educational systems are the successors of earlier more informal practices of social or cultural learning where children learnt from parents, other adults, and peers. Children were also inducted into relations of apprenticeship to expert craftspeople for the purpose of acquiring more specialized knowledge and skills. In these settings, there is little doubt that social learning aimed to induct children into a culture, including its ways of doing things, and its fund of collective knowledge (Sterelny 2012). Education of this sort was directly in the service of the flourishing of the community, rather than the individual,

even though the latter might benefit from membership in a flourishing community. The community's success depended on its ability to transmit its shared practical and theoretical knowledge to the next generation but also on its capacity to mould the young so that they acquired the same dispositions to act and believe of the adult members of the community. Conformism was prized because it facilitated the coordination of activities and the distribution of cognitive and manual labor. Individuals who think and act in similar ways can understand each other better than persons who have a different worldview (cf., Mameli 2001; Zawidzki 2018).

Two points emerge from thinking about cultural learning in pre-industrial settings. First, the primary epistemic aim was the epistemic well-being of the community rather than that of the individual learner. Second, learning was in the service of shaping the mind of the young so that they would come to resemble older members of the community. So, education encouraged the young to unquestioningly copy the mature members of the group. Many of the traits of character promoted by these forms of learning stand in stark opposition to the intellectual virtues mentioned by Croce and Pritchard. Intellectual servility, closed-mindedness, an unthinking tendency to imitate the behavior of older role models, would seem to be the kinds of disposition that educators sought to cultivate in the young. If this is right, open-mindedness, courage, and independence of thought would have been seen as obstacles to learning and consequently stifled.

One might react to these observations by accepting their historical plausibility and rejoicing in the fact that education no longer works that way. There is no doubt that modern educational principles and aims differ in several respects from earlier practices. But we understand these better if we appreciate the epistemic value of the model of education as induction of students into their culture by encouraging imitation, since this appreciation helps to see the underlying continuities of modern methods with pre-industrialized forms of education.

There can be epistemic value in the unthinking reproduction of cultural customs. This point is nicely defended by Levy and Alfano (2020) in their discussion of the transmission of cultural knowledge. They note, for example, that some communities might have accidentally stumbled upon procedures for the preparation and cooking of some food stuff such as cassava or corn that prevented poisoning or disease. The link between the product and ill-health is often causally obscure since not everyone gets sick, and the disease follows prolonged use and so it is not easily attributable to the food. In these circumstances, and without access to contemporary science, communities that flourished were those that somehow developed the correct procedures and had practices of knowledge transmission that involved the perfect imitation of customs.

Absolute, unthinking, imitation was essential as no member of the community knew exactly which aspects of the customary procedure were essential, and which were irrelevant, to disease prevention.

Reflection on the epistemic value of unthinking conformity to the customs of one's society reveals that an epistemic goal of current educational systems is to teach children to follow the social, moral, linguistic, and epistemic norms of the culture to which they belong. For example, language acquisition involves training children to use words in the way in which their community uses them. In this manner, children inherit a conceptual framework that embodies the epistemic resources available to them to make sense of their world. In short, to this day many educational practices aim to mould children to reproduce the behavior of adults. These practices rely on a tendency in human children to over-imitate the behavior of others (See Levy & Alfano, 2020 for a brief overview of the empirical evidence).

That said, there is little doubt that recent Western societies have come to value a special kind of individualism that is associated with the Enlightenment values of thinking for oneself. The intellectual virtues mentioned by Croce and Pritchard are those that are in line with this ethical and political outlook. Thinking for oneself requires courage, open-mindedness, and humility to avoid the twin risks of intellectual servility and hubris. Seen in the context of earlier approaches, this focus on these intellectual virtues is naturally understood as being, like its predecessors, a form of induction into a culture which, unlike some others, prizes innovation.

But if this is right, we have reason to believe that the primary epistemic goal of education is the epistemic well-being of the community. The strategies pursued to achieve this goal, even in industrialized societies, largely involve inducting students into the ways of speaking, thinking and acting of their community, and often depend on exploiting children's tendencies to imitate those who surround them. The relation of students to role models or exemplars would thus be one of sheer mimicking, rather than the reflective and selective emulative approach advocated by Croce and Pritchard.

That said, Western industrialized societies involve the creation of large-scale communities that generate new problems of coordination whilst largely depending on extensive specialization and division of cognitive labor. These societies, because of their size, can afford to foster competition between sub-groups trying to address the same practical problems so as to reap the epistemic benefit of tackling one problem in different ways. This novel approach to problem-solving requires a new kind of epistemic agent. It is in the service of this collective goal that contemporary education promotes the traits identified as virtues by Croce and Pritchard. And that is why education is a social activity with a collective epistemic goal.

References

- Levy, N., & Alfano, M. (2020). Knowledge from Vice: Deeply Social Epistemology. *Mind*, 129(515), 887–915. doi:10.1093/mind/fzz017.
- Mameli, M. (2001). Mindreading, Mindshaping, and Evolution. *Biology & Philosophy*, 16(5), 595–626. doi:10.1023/a:1012203830990.
- Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. Cambridge, MA: The MIT Press.
- Zawidzki, T. W. (2018). Mindshaping. In A. Newen, L. de Bruin, & S. Gallagher (Eds.), *The Oxford Handbook of 4E Cognition* (pp. 735–754). Oxford: Oxford University Press.

T&F Proofs – Not for Distribution

19c Commentary from Lani Watson

Croce and Pritchard's focus in this chapter is on the role of intellectual exemplars in intellectual character education. They align with the intellectual character education movement, which posits the cultivation of intellectual virtues as a key or primary goal of education, and ask; "how does thinking of the epistemic goals of education in this way inform our conception of education as an essentially social practice" (pp. 1–2)? In particular, they highlight "the extent to which social interactions with [intellectual] exemplars forms part of this educational method" (p. 2).

The chapter does valuable work in drawing attention to the as yet under-examined place of intellectual (as opposed to moral) exemplars in character education. It seems, for example, that there are good reasons to look more closely at the similarities and differences between moral and intellectual exemplarity as an educational tool, and Croce and Pritchard highlight some key features of this examination. Their conclusion with respect to the essentially social nature of education is perhaps somewhat too general to warrant in-depth critical engagement, but the implications of this conclusion for educational theory and practice are certainly worth exploring.

Firstly, one can ask what an emphasis on intellectual exemplars in education means for educational theory. In particular, Croce and Pritchard highlight the distinctive significance of direct, as opposed to indirect, exemplars, in the case of intellectual character education. This emphasis on direct interaction between students and exemplars underlines an already prominent feature of intellectual character education that pits it against (more traditional) educational theories premised on the mere transfer of epistemic goods via, for example, rote learning. This suggests, in turn, that theories of teaching and learning that likewise emphasize the value of direct instructional methods will sit well with an intellectual character education framing. Educational theories that foreground the value of close, relational interactions between teachers and students might fare particularly well. Moreover, emphasis on access to direct intellectual exemplars provides further theoretical support for initiatives that seek to reduce classroom sizes in primary and secondary education, or moderate

student numbers in line with available educators in institutes of higher education.

Secondly, one can ask what an emphasis on direct intellectual exemplars in education means for education practice. One basic practical implication is that teachers should themselves exhibit good intellectual character. Few, I think, would argue with this. Moreover, Croce and Pritchard are careful to emphasize that intellectual exemplarity need not come in the form of “intellectual sainthood”, absent any intellectual flaws, but is likely to be more effective in the form of “intellectual heroism”. The occasionally flawed intellectual hero is more relatable and, as such, “their very accessibility aids imitability” (p. 11).

This model of teachers as appropriately accessible, intellectual heroes is appealing. However, as always when it comes to translating rich theoretical insights into education practice, it seems likely that the devil is in the details. What degree of accessibility is appropriate for teachers and how best is it achieved? In particular, how does a teacher’s relatability as an intellectual exemplar intersect with their other roles and responsibilities as an educator, and how does or should it effect the power dynamics of the classroom? These questions point towards the need for further theoretical and empirical research on direct intellectual exemplars in education.

In addition, the focus on intellectual heroes, as opposed to other forms of intellectual exemplars, brings to the fore questions concerning the specific intellectual virtues that are or would be most valuable for teachers to exemplify. Are teachers better placed to exemplify intellectual courage or intellectual humility, for example, or should they (like the rest of us) aim for an appropriate mix of both. In general, how does a teacher’s role as an educator in the classroom (as opposed to her presence as a thinker in the world) influence the intellectual virtues that she should seek to exemplify, if at all.

One might wonder if there are, in fact, some intellectual virtues that teachers should actively avoid exercising. I have argued elsewhere, for example, that teachers should aim to ask fewer questions in the classroom, in order to cultivate the skill of good questioning (and intellectual virtuous character more generally) in students (Watson 2019). This, in turn, suggests that teachers should, at least sometimes, avoid exercising the intellectual virtue of inquisitiveness (and perhaps curiosity) in order to allow space for students to practice and refine these (in my view) essential and primary intellectual virtues (Watson 2016). The possibility that teachers may need to actively avoid exercising some intellectual virtues in the classroom raises, I think, the significance of attending to the details of the claim that intellectual exemplarity forms an essential part of the method of intellectual character education.

The questions I am raising here fall, to some extent, beyond the scope of the discussion presented by Croce and Pritchard. They are, at any

rate, not raised or explored in the chapter. Nonetheless, these questions do, I think, require careful consideration in order to move beyond the general conclusion that intellectual character education is essentially social, and towards an understanding of the theoretical and practical implications of this and, in particular, of placing a specific emphasis on direct intellectual exemplars as an educational tool.

References

- Watson, L. 2016. *Why Should We Educate for Inquisitiveness*. In *Intellectual Virtues and Education: Essays in Applied Virtue Epistemology* edited by Jason Baehr. New York: Routledge, 38–53.
- Watson, L. 2019. Educating for Inquisitiveness: A Case Against Exemplarism for Intellectual Character Education. *Journal of Moral Education* 48(3): 303–315.

19d Michel Croce and Duncan Pritchard's Response to Commentaries

The insightful commentaries to our contribution offered by Alessandra Tanesini and Lani Watson highlight some important aspects of the work that philosophers, education theorists, and educators should carry out to strengthen the theoretical and practical advantages of the educational approach we have proposed. The questions they raise would deserve a more in-depth treatment than we can offer here, but it seems important briefly to address a few points that could set the grounds for future investigations on intellectual virtue-based approaches to education and their social dimensions.

As Tanesini points out, our contribution is only a first step of a long journey into a social epistemology of education. From a truly social perspective—Tanesini argues—one might expect that the primary epistemic goal of education is the epistemic wellbeing of the community as opposed to the epistemic well-being of the individual learner. Her proposed comparison between the collective educational approaches of the pre-industrial settings and the ones we have adopted in Western industrialized societies reveals that imitation of exemplars or role-models plays a relevant role in both cultures albeit one that sacrifices individual flourishing to secure the wellbeing of the community at large.

This remark is helpful in that it helps us show that imitation of intellectual exemplars, at least in the educational model that we propose in the chapter, drastically departs from this common feature of the two cultures that Tanesini considers. A key feature of the virtue-based approach to education we favor is that it prevents the fostering of unquestioning attitudes in the learners via the emulation of mature members of one's community. Quite to the contrary, imitation of exemplars in an educational approach grounded in the social cultivation of intellectual virtue provides learners with the dispositions and intellectual resources to flourish as intellectual agents, and therefore as human beings.

In this respect, the educational approach we propose might in fact differ from the standard model of education in Western industrialized societies as we do not see the educational role of intellectual virtues as a way to induct learners into a particular culture, albeit one that makes room for innovation. Relatedly, we think that emulation of role models

does not—or, at least, should not—reduce to sheer mimicking. The view of education we have put forth is one in which the epistemic wellbeing of the community cannot do without ensuring that its members flourish at the individual level. Tanesini is right that the solution of problems arising in large-scale communities in Western industrialized societies requires a level of specialization and a cognitive division of labor that call for the exercise of intellectual virtues both at an individual and a collective level. Nonetheless, a social approach to education in a community need not forget about or sacrifice the intellectual flourishing of its members as a fundamental epistemic aim. For that would be too high a price to pay for proponents of the social epistemology of education.

That said, these considerations about the relationship between the goals of education at a collective and individual level can, at best, serve as a basis for future investigations into this issue. Two further questions raised by Watson add to this list: namely, what it means to emphasize the role of intellectual exemplars in education for both educational theory and educational practice.

As regards the former, Watson is surely right that educational theories that promote relational interactions between teachers and learners sit well within the approach we have proposed. It is also true that this might speak in favor of reducing the students–teachers ratio where it is possible. Yet, the fact that we highlight the role of direct exemplars in intellectual character education should not lead one to think that our approach is necessarily bound to direct instructional methods, if what we mean by that is an old-fashioned teaching approach based on taught classes delivered by the teacher. It is surely key to an intellectual virtue-based approach to education that the students learn what the virtues are, how they work, and why we need them. But prominent examples like the *Intellectual Virtues Academy* (Long Beach, California), the *Anteater Virtues Project* (University of California, Irvine), and the education programs that the *Institute for the Study of Human Flourishing* (University of Oklahoma) conducts with schools in the Oklahoma City area reveal that character education is compatible with a wide variety of strategies to help students learn and develop moral, intellectual, and civic virtues.¹

As regards the latter question, it is surely helpful to think about which intellectual virtues the teachers should exercise and which virtues, as Watson argues, they should avoid exercising for the epistemic good of their students. Allowing space for them to become inquisitive epistemic agents might well require that the teacher refrains from modeling the virtues of good questioning and curiosity. This interesting case allows us to highlight the importance of involving the students themselves in the role-modeling phase of virtue development. Rather than indulging in an explanation or an exemplification of how an inquisitive person asks questions, the teacher would be particularly helpful in highlighting those episodes in which some students, with their questioning or curious

attitudes, can serve as intellectual exemplars for their classmates. No matter what set of intellectual virtues proves to be fundamental for teachers *qua* role models, the above considerations reveal that teachers cannot do without one of the most complex virtues, namely practical wisdom. For in most cases, the choice between various possible strategies to help the students develop their virtues will be determined by peculiar features of the classroom and the social environment. Practical wisdom allows the teacher to tailor their teaching strategies to the overall situation of their class and coordinate the joint activity of several intellectual virtues that are key to maintaining such a dynamic approach. Interestingly enough, then, teachers cannot but practice an overarching intellectual virtue that the students will be able to recognize only once they have developed several other intellectual virtuous traits.

Note

- 1 For a recent educational study of the Anteatler Virtues project at the University of California, Irvine, which also summarizes the pedagogical strategies employed and attempts to measure their effects on student learning outcomes, see Orona and Pritchard (2021).

Reference

- Orona, G. A., & Pritchard, D. H. (2021). 'Inculcating Curiosity: Pilot Results of an Online Module to Enhance Undergraduate Intellectual Virtue', *Assessment & Evaluation in Higher Education*, doi: 10.1080/02602938.2021.1919988.

T&F Proofs – Not for Distribution