

*Routledge Studies in Epistemology*

**TRUST RESPONSIBLY**  
**NON-EVIDENTIAL VIRTUE EPISTEMOLOGY**

Jakob Ohlhorst



‘Ohlhorst sets forth a novel virtue-theoretic solution to a basic problem of hinge epistemology, the problem of distinguishing good from bad hinges. The book is a must read for anyone interested in hinge epistemology.’

**Sven Bernecker**, *University of Cologne, Germany*

‘Ohlhorst’s book masterfully merges hinge (or non-evidential) epistemology and a dual theory of intellectual virtues. The result is an original, highly interesting view that sheds new light on debates concerning scepticism, hinge epistemology, and virtue epistemology. I strongly recommend reading the book. It is an excellent piece of work.’

**Nikolaj Pedersen**, *Yonsei University, South Korea*

‘This is an elegantly written, accessible, and sharply argued book that uses the resources of virtue epistemology to defend hinge epistemology. A substantial and eye-opening plea for hinges that might move even staunch skeptics about hinges.’

**Thomas Grundmann**, *University of Cologne, Germany*



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

# Trust Responsibly

This book offers a defence of Wrightian epistemic entitlement, one of the most prominent approaches to hinge epistemology. It also systematically explores the connections between virtue epistemology and hinge epistemology.

According to hinge epistemology, any human belief set is built within and upon a framework of pre-evidential propositions – hinges – that cannot be justified. Epistemic entitlement argues that we are entitled to trust our hinges. But there remains a problem. Entitlement is inherently unconstrained and arbitrary: We can be entitled to any hinge proposition under the right circumstances. In this book, the author argues that we need a non-arbitrariness clause that protects entitlement from defeat. This clause, he argues, is to require epistemic virtue. Virtuous cognitive dispositions provide the non-arbitrariness clause that protects entitlement from defeat. The epistemic character of the agent who holds a particular set of hinges tells us something about the hinges' epistemic status. Conversely, epistemic virtues are cognitive dispositions and capacities that rely on hinge propositions – without trusting in some hinges, we would be unable to exercise our virtues.

*Trust Responsibly* will appeal to scholars and advanced students working on epistemology, Wittgenstein, and virtues.

**Jakob Ohlhorst** is a postdoc at the University of Vienna and VU Amsterdam. Previously, he completed his Dr. Phil. at the University of Cologne. He has published 'Epistemic Austerity' at *Synthese* and 'Dual Processes, Dual Virtues' at *Philosophical Studies*; both papers make key arguments for this book.

**Routledge Studies in Epistemology**

Edited by Kevin McCain

*University of Alabama at Birmingham, USA*

and

Scott Stapleford

*St. Thomas University, Canada*

**Epistemic Care**

Vulnerability, Inquiry, and Social Epistemology

*Casey Rebecca Johnson*

**The Epistemology of Modality and Philosophical Methodology**

*Edited by Anand Jayprakash Vaidya and Duško Prelević*

**Rational Understanding**

From Explanation to Knowledge

*Miloud Belkoniene*

**Illuminating Errors**

New Essays on Knowledge from Non-Knowledge

*Edited by Rodrigo Borges and Ian Schnee*

**Digital Knowledge**

A Philosophical Investigation

*J. Adam Carter*

**Seemings and the Foundations of Justification**

A Defense of Phenomenal Conservatism

*Blake McAllister*

**Trust Responsibly**

Non-Evidential Virtue Epistemology

*Jakob Ohlhorst*

For more information about this series, please visit: [www.routledge.com/Routledge-Studies-in-Epistemology/book-series/RSIE](http://www.routledge.com/Routledge-Studies-in-Epistemology/book-series/RSIE)

# Trust Responsibly

## Non-Evidential Virtue Epistemology

Jakob Ohlhorst

First published 2024  
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*

© 2024 Jakob Ohlhorst

The right of Jakob Ohlhorst to be identified as author of this work 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.

ISBN: 978-1-032-46098-7 (hbk)

ISBN: 978-1-032-44907-4 (pbk)

ISBN: 978-1-003-37446-6 (ebk)

DOI: 10.4324/9781003374466

Typeset in Sabon LT Pro  
by Apex CoVantage, LLC

Für Susanne und Nils





**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

# Contents

<i>Acknowledgements</i>	<i>x</i>
Introduction	1
1 Hinges, certainties, and trust	5
2 Entitlement	35
3 Problems for entitlement: demarcation, arbitrariness, and relativism	73
4 Virtue	90
5 Trust virtuously	114
Conclusion	141
<i>Index</i>	<i>145</i>

# Acknowledgements

I would like to thank my thesis supervisors, Sven Bernecker and Nikolaj Pedersen, for their help and advice during the redaction of my thesis. Additionally, I would like to thank the supervisors who supported me during my research stays in Irvine and Glasgow: Duncan Pritchard, Annalisa Coliva, Adam Carter, and Jack Lyons.

I am very grateful to everyone who helped me with advice during the writing of this thesis. Notably, Philip Ebert and Lisa Bastian played a seminal role for this book. While completeness is hard to achieve, I want to particularly thank Anne Meylan, Thomas Grundmann, Lisa Benossi, Amy Flowerree, Steffen Koch, Luis Rosa, Paul Silva, Eyal Tal, Peter Graham, Francesco Praolini, Giulia Napolitano, Sofia Bokros, Dominik Balg, Eve Kitsik, Chris Ranalli, Taylor Dunn, Oscar Piedrahita, Martin Miragoli, Aengus Daly, and Sophia Fazio.

This book was written with support from the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research, and the Swiss National Science Foundation. It received language editing, thanks to the Katharina Niemeyer-Grant of the a.r.t.e.s Graduate School at the University of Cologne. The open access publication of this book has been published with the support of the Swiss National Science Foundation.

This work is a doctoral dissertation accepted by the Faculty of Arts and Humanities of the University of Cologne.

# Introduction

*Strange coincidence, that every man whose skull has been opened had a brain!*  
(Wittgenstein, 1969, §207)

Ludwig Wittgenstein's joke points to a strange feature of our epistemology. Even though very few of us have ever seen a brain and even though most brains will never be seen, we are certain that everyone has a brain. To ruin a perfectly good joke, we simply take the existence of brains in our skulls for granted and do not think it is a strange coincidence that the skulls we have actually opened did contain brains.

The certainty that everyone has a brain is a nice example of a kind of belief that plays a particular role in our epistemology. These certainties lie outside the normal push and pull of evidence or justification and consequently they have been largely overlooked by mainstream epistemology, presumably because of the assumption that once we fix the other epistemological problems, these certainties will end up looking just like normal beliefs.<sup>1</sup>

I prefer a different approach: *Hinge epistemology* proposes a sophisticated account of these certainties. It is strongly influenced by Wittgenstein's *On Certainty* (1969), which uses the term 'hinge belief' for these peculiar certainties. In Chapter 1, I introduce my own view of these Wittgensteinian certainties. I argue that in any belief system there have to be some propositions that we cannot support in a non-circular fashion. I call these propositions 'hinges.' I further argue that, psychologically speaking, we are certain of these hinges but, given their hinge status, we have to *trust* that they are indeed the case.

Some hinge epistemologists take hinges as being entirely beyond the reach of epistemology. They argue that hinge propositions cannot be true and that consequently they also cannot be epistemically evaluated (Moyal-Sharrock, 2004). While agreeing that hinge certainties cannot be justified like regular beliefs, I nonetheless reject this approach. We can get trusting in hinges right or wrong. Following Crispin Wright (2004), I argue in

## 2 Introduction

Chapter 2 that we are entitled to trust in hinge propositions because they enable cognitive activity. I also explain how entitlement works as a solution to the sceptical problem.

This debate plays out against the background of the debate about epistemological internalism and externalism. Hinges, which are commitments arising from our regular beliefs, as well as Wrightean entitlement, are internalist notions (Pritchard, 2016). Given their pre-evidential status, which lies beyond standard epistemic evaluation, the problem of the criterion (Chisholm, 1973) becomes particularly acute with regard to hinges and entitlement. What criteria distinguish epistemically good from epistemically bad hinges? What are the constraints for entitlement such that it can fulfil its epistemological function? In Chapter 3, I explain exactly how this problem arises. It is a product of the *arbitrariness* of hinge certainties. The demarcation problem, as Wright (2004) calls it, also explains social epistemological problems such as epistemic relativism and the phenomenon of irresolvable deep disagreement.

My solution to this arbitrariness problem is to shift the evaluation from the agent's doxastic states to the epistemic agent herself. To this end, I develop my own account of *epistemic virtues* in Chapter 4. Appealing to *dual process theories* from cognitive psychology, I argue that the competing accounts of epistemic virtue – *virtue reliabilism* and *virtue responsibilism* – are complementary. Each kind of virtue corresponds to a type of cognitive processing. This account has the additional benefit of naturalising our epistemic virtues.

Finally, I explain how this account of epistemic virtue can solve the problems that arise from the arbitrariness of entitlement. I do this by arguing that cognitive activity is an epistemic good and epistemically virtuous cognitive activity is an exemplary epistemic good. In order for our virtues to function, however, we have to presuppose some hinges. We therefore have a non-arbitrary entitlement to trust the hinges that we need for our virtues to be epistemically efficacious.

The chapters mutually support one another, relying on arguments previously made, but they are also more or less self-contained. Only Chapters 3 and 5 substantially depend on what is said in the preceding chapters.

The dispute between epistemic internalism and externalism derives its force from the fact that good epistemic standing can be taken to mean two different things. One is objective *de facto* epistemic success in terms of true beliefs, and the other is subjective good epistemic conduct from the limited perspective of the agent (Goldman, 1988). The two notions have resisted being reduced to each other.

In this work, I propose an account that explains why we need both views and how they fit together. We accordingly need to look at our most fundamental convictions, that is, hinge certainties, because both internalist

justification and standard externalist justification break down at that level. Neither type of justification can be gained for the certainty that we have a brain or some more fundamental hinge certainties, for example, that others have a mind.

Nevertheless, these fundamental commitments can be warranted – and this warrant is called entitlement. My argument is that we are entitled to trust that these hinges are true because this trust enables the very virtues that we need to gain either externalist or internalist justification for our other beliefs. There are two types of cognitive processing. Type 1 is fast and automatic, and Type 2 is slow and controlled. Virtues of Type 1 generate externalist justification and virtues of Type 2 generate internalist justification and other epistemic values. These two virtue types, however, require trust in certain hinges. We are entitled to trust in these hinges because this is the transcendental condition for exercising these virtues.

This solution requires that I thread the needle between the requirements of internalism and externalism. On the one hand, the internalist warrant that is entitlement is subject to externalist criteria and, on the other, externalist warrants are supplemented by internalist requirements. A key part of my response is the claim that warranting at the fundamental level of entitlement works following different rules than the rules at the level of justification and knowledge on which the internalism–externalism debate usually plays out.

This focus on hinges also has consequences for how we do philosophy. Philosophical investigation is about our most fundamental beliefs. It examines them and studies their epistemic status. However, this takes philosophical investigation out of the safe waters of regular epistemic justification. I believe that this explains the lack of convergence and great diversity of philosophical research (Chalmers, 2015).

## Note

1. Let us bracket for now the testimony of neurosurgeons that is available to us. Many of Wittgenstein's examples have aged badly, but I am unwilling to let this fact ruin a colourful illustration. It still applies to our everyday epistemology, as few of us ever deal with neurologists.

## References

- Chalmers, D. J. (2015) 'Why Isn't There More Progress in Philosophy?', *Philosophy*, 90(1), pp. 3–31. <https://doi.org/10.1017/S0031819114000436>.
- Chisholm, R. M. (1973) 'The Problem of the Criterion', in *The Foundations of Knowing*. Minneapolis: University of Minnesota Press, pp. 61–75.
- Goldman, A. I. (1988) 'Strong and Weak Justification', *Philosophical Perspectives*, 2, pp. 51–69.

#### 4 Introduction

- Moyal-Sharrock, D. (2004) *Understanding Wittgenstein's on Certainty*. Basingstoke: Palgrave Macmillan.
- Pritchard, D. (2016) *Epistemic Angst*. Princeton: Princeton University Press.
- Wittgenstein, L. (1969) *Über Gewissheit – On Certainty*. Edited by G. E. M. Anscombe and G. H. von Wright. Oxford: Blackwell.
- Wright, C. (2004) 'I – Crispin Wright: Warrant for Nothing (and Foundations for Free)?', *Aristotelian Society Supplementary Volume*, 78(1), pp. 167–212. <https://doi.org/10.1111/j.0309-7013.2004.00121.x>.

# 1 Hinges, certainties, and trust

## 1.1 Introduction

In this chapter, I introduce three related notions: hinge, certainty, and trust. Each of these terms describes a peculiar feature of single beliefs and how they relate to other beliefs. These features are usually overlooked in epistemology because they are purely doxastic. They do not give any information about their own truth or falsehood. Instead, they describe psychological and doxastic mechanisms inside a subject's belief system. The external facts are bracketed.<sup>1</sup> These concepts play an important role in internalist epistemology. In this chapter, I simply describe how I believe that agents structure their belief systems without taking any position on the beliefs' epistemic status.

A belief in a *hinge proposition* is a special kind of certainty. Hinge propositions are characterised by the fact that their content determines how other propositions relate to each other. They do so either as a *cornerstone* or as a *rule*. On the one hand, cornerstone beliefs determine how the concepts in other beliefs relate to each other; on the other hand, rules determine how different beliefs support each other as reasons. To do this stably, they must be certainties, otherwise whole swaths of our beliefs would permanently shift in content or have to be continuously re-evaluated.

A *certainty* is a belief that is held with such a high degree of confidence that other beliefs or further evidence do not influence its doxastic status. Not every certainty is about a hinge. Once you are certain of something, it stands firm – no epistemic mechanism will either strengthen or weaken your confidence in it.<sup>2</sup> Evidence and other beliefs become moot for certainties.

I call *trust* a doxastic attitude that is adopted without requiring any sufficient evidence for its content. Hence, in any case where you accept something to be the case without having the necessary evidence at hand, you trust it to be the case. I claim that accepting hinges involves a special kind of trust.<sup>3</sup>



## 6 Hinges, certainties, and trust

I have just used a very wide notion of ‘belief.’ On this count, believing a proposition simply means endorsing the proposition as being the case (Schwitzgebel, 2019). This notion tracks the ordinary language use of ‘belief.’ Many hinge epistemologists reject the idea that hinges can be believed. Crispin Wright (2004, pp. 176–177) suggests that we ‘accept’ hinges because some epistemologists consider belief to be essentially bound by evidence.<sup>4</sup> Annalisa Coliva (2015, p. 44) thinks that hinges are rules and therefore cannot be believed. Also, Duncan Pritchard (2016, p. 92) calls the relevant doxastic state ‘commitment’ instead of ‘belief’ because he does not think that hinges are candidates for knowledge.

I agree with Pritchard that we are committed to our hinges, and that our hinge commitments are no ordinary beliefs. Nevertheless, I think that our commitment to hinges is the endorsement of something as true or accurate – even rules can be accurate or not. I played Monopoly with the wrong rules for most of my life. Like Wittgenstein (1969, §§159, 253), I will follow ordinary language which also calls doxastic states that are insensitive to evidence or no candidates for knowledge ‘beliefs.’ For instance, a religious fanatic can be described as having religious beliefs even though these are insensitive to evidence and not candidates for knowledge. Also, ideologies and delusions are frequently described as ‘beliefs.’ I will treat belief as the genus, hinge beliefs and (evidence-sensitive, knowledge-apt) ordinary beliefs will be the species. For a careful debate of these issues, see Moyal-Sharrock (2004, pp. 181–184) and Lopez Barranco (2023).

I first describe how believing and accepting hinge propositions work. Second, I introduce and illustrate the notion of certainties using a diverse set of examples. Third, I explain why our belief in hinges is certain. Finally, I examine the notion of trust, in particular, the *epistemic trust* that we have in hinges.

### 1.2 Hinges

Within our belief systems, there are necessarily some *hinge propositions* that are necessary for the functioning of our believing and acting. I introduce the notions of hinge, rule, and cornerstone propositions. I use ‘hinge’ to group both rules and cornerstones: Every hinge proposition is either a rule or a cornerstone proposition. So what is a hinge? According to Wittgenstein, these are the propositions on which our whole world view ‘turns’ (Wittgenstein, 1969, §341). I will spell out this metaphor in this section.

#### 1.2.1 Rule and cornerstone propositions

Rules and cornerstones are propositions in our belief systems that relate to other believed propositions in a special way. Namely, they have many

implications for what and how we believe. I will begin with the following definitions of a cornerstone and a rule proposition:

- (CS) A proposition functions as a *cornerstone proposition* within a belief set if and only if the proposition is a necessary condition for a large set of other believed propositions. Cornerstones thereby determine the content of other propositions.

‘Everybody has parents’ is a Wittgensteinian example of a cornerstone. It is a presupposition for essentially any biographical fact we could think of. No one just popped into existence. For someone to have done something or been somewhere, they first had to be born to a father and a mother (Wittgenstein, 1969, §211). ‘Everybody has parents’ thereby determines the content of our beliefs about other people.

- (R) A proposition functions as a *rule proposition* within a belief set if and only if the proposition expresses a rule that is a necessary presupposition for rationally changing our belief set or rationally acting on the basis of our belief set. Rules determine how different believed propositions relate to each other as reasons.

‘Timetables are a reliable source of information about train departures’ is a fairly undemanding rule proposition. It is presupposed anytime we plan a trip or form other train-travel-related beliefs. The proposition expresses the rule that for this kind of information, I can rely on that kind of source of evidence. Rule propositions determine what counts as evidence. Coliva (2015) puts this notion of a rule at the centre of her hinge epistemology and Goldberg (2015, p. 282) calls these propositions, that we are entitled to accept, *KR-Propositions*.

Our belief system is highly complex. Quine and Ullian (1978) called it a ‘Web of Belief.’ The metaphor of the web refers to beliefs cohering in all sorts of ways. For example, an individual’s belief system is structured through semantic associations. Our beliefs about cats tend to evoke beliefs about dogs, because the two animals are semantically closely associated as animals who are the paradigms of pets.

For the purpose of discussing cornerstones, however, I will ignore the many kinds of connections and simply focus on entailment relations. Each belief you hold can be considered as containing one proposition, which means your belief set corresponds to a set of propositions.

There are hosts of entailment relations between the singular propositions we believe: Some propositions entail each other mutually, others form long chains, still others are fairly isolated, not implying much. You do not need to know all those entailment relations; indeed you most certainly do not

know all entailments, you are not logically omniscient. However, a considerable subset of these entailment relations will be accessible to you if you consider them. Consider the belief ‘my tea has gone cold.’ This implies that ‘the tea once was hot,’ which further implies that ‘objects can change their temperature’ and so on.

Arguably, everybody’s belief system has a somewhat similar structure. Nobody, for example, has a totally ordered chain of believed propositions where there is no symmetry of entailment between any two propositions. Rather, there are peripheral and more central propositions in our belief set. A proposition is peripheral if and only if it is entailed by no other propositions. A proposition is more central if it is entailed by other propositions; that is, it is a necessary condition for them. A proposition will be maximally central if it is entailed by all other propositions in a belief set. It will be maximally peripheral or minimally central if it is entailed by no proposition. Hence, centrality is a graded notion. Logical necessities are an example of maximally central propositions, while perhaps a statement about the charge of some particular particle could be minimally central (*cf.* Quine and Ullian, 1978).

Further, it can be argued that a belief system is divided into subdomains. There are peripheral propositions entailing central propositions, but these peripheral propositions do not entail other equally or even more central propositions. There are, for example, fairly few entailment relations between propositions about architecture and propositions about biology. A peripheral architectural proposition about an Ottoman mosque, for example, about the number of tiles in its dome, will have many implications for central propositions about architecture and geometry, but arguably not a single implication for even very central biological propositions about sloths.

Hence, holding a peripheral proposition commits you to accepting more central propositions too, because the former entails the latter. *Vice versa*, if you reject a central proposition C, then you cannot rationally hold peripheral propositions for whose truth the central proposition is a necessary condition. Crispin Wright (2004, pp. 167–168) calls these central propositions cornerstones. However, his definition uses warrant as the criterion for the cornerstone status, instead of accepted truth. According to Wright, a cornerstone’s warrant is a necessary precondition for the warrant of more peripheral propositions. Under the assumption that warrant transmits just like (believed) truth across known entailments, the two criteria become equivalent. These cornerstones also form a subset of the more broad notion of hinges (Coliva and Moyal-Sharrock, 2016).

The entailment relations of the members of the set of propositions that make up your belief set thereby force you to accept other propositions. Notably, your belief set forces you to accept or be committed to the central

cornerstone propositions. Hence, in my book, ‘cornerstone’ is a semantic concept that determines the structure of someone’s belief set.

Although the cornerstones are entailed by more peripheral propositions, they conversely determine the content of the peripheral propositions. Peripheral propositions’ content is made up of elements characterised and defined by the cornerstones. Without the cornerstones, the peripheral beliefs would be meaningless. Peripheral beliefs about particular bodies would lack in content without the more central beliefs setting out what a body is.

This structural fact is the first way in which we are committed to some hinges; that is, simply by virtue of accepting a set of propositions. It also clearly demarcates the class of cornerstones as containing those beliefs that are entailed by many propositions. It is further important to note that not everyone will hold the same propositions as being central. The overall sets of accepted propositions may vary considerably from individual to individual, which means that they will always entail slightly different central propositions and cornerstones. The overall sets of accepted propositions in different epistemic communities may be so different that the cornerstones could just have a minimal overlap. Consider the deeply divergent metaphysical presuppositions of a monistic theist who believes that reality derives from the action of a single god versus the presuppositions of a materialist atheist who takes matter to be all there is.

The second kind of commitment to hinges comes from rules. Rule propositions work a bit differently. They are not just entailed by a static set of believed propositions. Rule propositions are the propositions that we are committed to when we rationally *change* or *use* the set of accepted propositions.

To update the set of accepted or believed propositions or to act on its basis requires a set of rules. These rules can be formulated as propositions. I call these propositions ‘rule propositions.’ As cornerstone propositions are entailed by the accepted set of propositions, rule propositions are entailed by the changes we make to the set of accepted propositions or by the practical applications we draw from out of the set of accepted propositions. Note that, like cornerstone propositions, rule propositions may not be explicitly believed, but they are entailed by how the agent changes and uses the set of believed propositions. In this way, rule propositions become members of the set of propositions to which the agent is committed (Coliva, 2015).

The claim that rules cannot be identical to the propositions expressing them has been much debated (Carroll, 1895). Nevertheless, if we follow a rule, then we are arguably also committed to accepting the proposition that expresses the rule.<sup>5</sup> Thus, this rule proposition is a necessary doxastic precondition for changes to the set of propositions accepted on the basis of the rule. If we rejected the rule proposition, then we would be irrational in

changing the set of accepted propositions in accordance with the rule. The rule functions as an implicit premise. Hence, changing the set of accepted propositions using rules also commits us to rule propositions – and we clearly do follow rules.

Rule propositions are the propositional correlates of the rules we apply when we take something to be a reason for something else. Consider, for example, the well-studied rule of *object permanence*: ‘objects that have been occluded from view are still there’ (Carey, 2009, p. 40; Greco, 2021, p. 118). This is the corresponding rule proposition to which we are committed. The rule of object permanence partially determines the evidential role that, for example, a memory about where I put my copy of *On Certainty* plays for my beliefs about its current location. Objects, and therefore books, do not just disappear. Therefore, *On Certainty* is still where I remember putting it. Another rule at play here is that I can rely on distinct memory impressions.

Hence, hinge propositions determine what counts as evidence for what: On the one hand, because cornerstone propositions determine semantic relations and on the other hand, because rule propositions determine what counts as a reason for what. There is some overlap between cornerstones and rules. By entailment, peripheral propositions are evidence for more central propositions, while the central propositions lay down the rules for how peripheral propositions imply other propositions. Some cornerstones are therefore also rule propositions that lay down some of the ways in which the set of accepted propositions can be expanded or reduced.

Conversely, there are some rule propositions that determine other propositions’ semantic content. Consider again the example ‘I have a body.’ This rule also is a cornerstone insofar as it is implied by a wide range of propositions about me and I need to accept it as true to be able to believe and even understand beliefs about, for example, my hands.

This overlap is unsurprising because cornerstones and rules were originally proposed as competing views of hinges. Wright (2004) argues for cornerstones based on entailment, while Coliva (2015) defends rules that characterise our (epistemic) practices. Both of these accounts can, to a certain degree, incorporate the other, but I would argue that the distinction is useful to emphasise the different ways how a proposition can be a hinge.

Finally, in acting, we are implicitly committed to certain rule propositions. For instance, in order to intentionally throw a ball, you need to implicitly accept that the ball will behave in a certain way. You need to accept, among other things, that it is inert and does not change its momentum without external influence. Thus, each action and your conceiving of yourself as acting has necessary preconditions. If you rejected these propositions, you would be unable to conceive of yourself as acting.

If we follow Anscombe (1957, p. 13), who argues that action always involves knowledge of what you are doing, then you will be unable to act without accepting some hinge propositions. Knowledge of your action means that it is part of the set of accepted propositions. Accepting the proposition ‘I am x-ing’ will then entail a range of hinges to which you are committed by your acceptance and consequently by your action. Acting therefore commits us to hinges by simple entailment.

### 1.2.2 Hinges and contemporary epistemology

The account of hinges developed here is an account of *narrow mental content* (Brown, 2022). It is a theory about the structure of individuals’ belief sets and how this explains their behaviour that only relies on factors internal to the respective individuals. More specifically, it is a so-called *conceptual role semantics* (Block, 1998). Our hinges determine the content of and evidential relations between our other beliefs – content is defined through a concept’s role in our belief system. Conceptual role semantics appeal to the Wittgensteinian adage that *meaning is use*.

The term semantics is liable to lead to confusion in this case because, typically, semantic content is characterised by when a proposition is true and when it is false – semantics is often truth-functional. My account is incomplete in this respect – it does not aim to explain when a proposition is true and when not, only how the subject relates different propositions and inputs. What is missing in this work is an account of *wide content* which gives the semantics truth-functionality. Consequently, this is part of a *two-factor theory of meaning*, as Ned Block (1998) would call it.

Some would deny that this fragment of a semantic model should be called an account of content because it lacks the relevant connection to the world. I grant this missing connection, still it is an account of content insofar as it explains the structure of the agent’s understanding and beliefs. One way of spelling out an account of narrow content has been undertaken by David Chalmers (2003). He describes our narrow mental content in terms of scenarios that we accept or exclude. Scenarios can be described as possible worlds. In Chalmers’s terminology, our hinges would demarcate the set of epistemically possible worlds for an agent from the epistemically impossible worlds.

The status of narrow mental content is controversial (Brown, 2022). A prominent recent challenge has been raised by Yli-Vakkuri and Hawthorne (2018), who argue that any meaningful conception of narrow content will lead to a problematic relativism. Given that my account is incomplete in this respect, I will remain silent on threats of *semantic* relativism, but as I show in Chapter 3, epistemic relativism is baked into the foundations of hinge epistemology.

A related issue that arises from my account of hinges as determining narrow content is that having a hinge framework does not guarantee successful communication through shared hinges. Many hinge epistemologists appeal to Wittgenstein's considerations in the *Philosophical Investigations* (Wittgenstein, 1958) to argue that hinges can only function as such if they are shared in the linguistic community – see, for example, Moyal-Sharrock's (2004) concept of *objective certainties*. On these views, hinges fix agent-external wide content.

I agree that hinges need to be shared to a certain extent in order to enable linguistic exchange, and if two agents are able to communicate, this is explained by their sharing a significant portion of their hinges. However, we have to distinguish the doxastic function of hinges from their linguistic function. The latter can be exercised without the former.<sup>6</sup> Hinges sometimes diverge, which explains the phenomena of irresolvable deep disagreement (Ranalli, 2020) or the ununderstandability of delusions (Ohlhorst, 2021).

An important feature of hinges, that is, rules and cornerstones, is their peculiar evidential situation. Whatever you believe, start believing, or stop believing, you always already have to presuppose some cornerstones and rules to do so. You need to presuppose the cornerstones for the beliefs to even have narrow semantic content, that is, to mean something, and you need the rules to rationally adopt and support the beliefs. This means that your hinges cannot be supported by evidence without already being presupposed. Hence, any evidential support for the rules and cornerstones that you accept is *necessarily circular*. In this chapter, I am however more concerned with how we take our beliefs to be related rather than their epistemic status and whether hinges are justified or known.

I will dive deeper into the epistemic consequences of this circularity in Chapters 2 and 3, but I want to give some pointers on how this view integrates with the broader epistemological debate. Hinges are what Fred Dretske (2005) calls 'heavyweight propositions.' These are propositions that are entailed by our ordinary evidence-based beliefs – for example, the ordinary belief that this tree stump has 300 tree rings and therefore is 300 years old entails the heavyweight proposition that the world is older than you. But because heavyweight propositions are *de facto* presupposed as implicit premises in this entailment, the heavyweight proposition is only circularly supported. This way, hinges or heavyweight propositions end up serving as evidence for themselves which is a vicious circularity.

You might object that we nevertheless can gain non-vicious circular justification, namely by establishing a track record of our rules. An early version of this defence of circularity was developed by Braithwaite (1968, pp. 276–277), arguing that we can establish the validity of a rule by using this very rule – while presupposing a cornerstone would indeed be viciously circular. However, gaining circular justification through a track record is not as easy as it might seem. Note that the famous track

record argument with Roxanne relying exclusively on her gas gauge and no background knowledge to establish the gauge's reliability was proposed by Vogel (2000) as a paradigmatic example of illegitimate bootstrapping which challenges circular arguments.

The defenders of circular track record arguments take two avenues. The first avenue is to argue that the circularity of track record arguments is unproblematic if we grant that we have foundational justification for the beliefs that establish our track record (Bergmann, 2004, p. 723; Schmitt, 2004, p. 395). Note, however, that a hinge epistemologist would argue that this foundational justification already presupposes rule propositions as implicit premises – if the foundationalist denied the rule she would undermine her foundational justification (*see also* Lynch and Silva, 2016). I will examine foundational sources of justification like reliabilism and evidentialism in more detail in Chapter 2.

The second argument in favour of track record arguments is that they are the only way to avoid scepticism. More exactly, Van Cleve (2003) defends the circular track record argument by arguing that the alternatives – coherentism, common sense Reidianism, or Sosa's two-level account<sup>7</sup> – are equally circular as the track record argument. Consequently, we have to endorse the circularity in order to stave off scepticism.

Hinge epistemology, however, is a fourth, non-sceptical alternative. Ten years after his defence of the track record argument (2003), Van Cleve distinguishes (2013, p. 256) four ways how our beliefs can be structured as an alternative to scepticism.<sup>8</sup> A. We can be *infinitists* and accept that every belief we have is supported by an infinite non-circular chain of further beliefs. B. We can be *coherentists*, accepting that any evidence in the long run ends up being circular. I will raise my misgivings about coherentism and infinitism in Chapters 2 and 3. C. Most authors defend some type of *foundationalism*, arguing that certain beliefs have a weak or strong *prima facie* justification. As said, I will consider foundationalism more closely in Chapter 2. This leaves the oft-overlooked fourth option which I endorse, D. *Positism*. Hinge certainties are the unsupported presuppositions on which our other beliefs are built.

By playing this fundamental role, our hinges also demarcate the limits of our inquiry (Wittgenstein, 1969, §204). Given that we cannot find independent evidence for them and that there are no meta-rules establishing how the rules of investigation should be developed, we cannot fruitfully investigate the nature of hinges. The only thing we would be doing is pushing the boundary back to a further hinge without strengthening our epistemic position. Similarly, refuting hinges does not make them go away, they simply get superseded by other rules and cornerstones (Carey, 2009, p. 87).



### 1.3 Certainty

I define certainty as follows:

- (C) A belief that P is a *certainty* if and only if it is held with such a degree of conviction that no other belief or evidence would change the degree of conviction with which it is held.

Assuming that belief change is only epistemically rational if there is some evidence I recognise as defeating the belief, it follows that certainties cannot be rationally abandoned. Certainties are the beliefs that stand firm. Note also that not every certainty needs to be about a hinge.

Note that this is a mere *psychological certainty*. Nothing the world could throw at me could bring me off course through rational mechanisms, and as a matter of anthropological fact we do not abandon our certainties.<sup>9</sup> I here remain strictly neutral about how reasonable such a stance is as I am only claiming that we can and sometimes do have such beliefs. Consequently, certainties are convictions which we cannot rationally doubt. If you are certain of P, there is no sort of evidence that could make you genuinely doubt it; if you were to start doubting P, this would be an arational or even irrational process not underwritten by any reasons accessible to you.<sup>10</sup>

Note that, genetically speaking, there may be two kinds of certainties. We can call these unconditional certainties and conditional certainties, respectively. The former are acceptances or beliefs certain for a subject independently of the evidence, for example, tautologies or also *hinge certainties*, as I will argue in Section 1.4. The others have not always been certain, but the agent has acquired so much independent evidence for them that they have become certain.

You might think that this is not possible. But it is not as exotic as it might sound. In Bayesian epistemology, certainty corresponds to credence of 1, which can never be undermined.<sup>11</sup> In modal epistemology, this would correspond to a strong necessity operator scoping over all accessible worlds or epistemic possibilities such that there is no doxastically accessible case where the belief does not hold. In other words, certainties are at least a natural possibility in formalisations of our belief systems (Beddor, 2020, pp. 10–11).

In folk epistemological terms, certainty could just be a belief that stands firm for you. Nothing anyone tells or shows you could dislodge it. You *cannot seriously entertain* any evidence whatsoever that could dislodge it because you lack the background beliefs and conceptual resources that would allow you to conceive of or grasp any such evidence. Take, for example, the idea that you were wrong in all your simple arithmetic calculations until now: What sort of evidence could support such a belief? (Wittgenstein, 1969, §55).

I would further argue that certainties, once they are certain, do not rely on evidential support. Once you are certain of something, it is not just beyond any defeating evidence, but beyond all evidence. You are so convinced of its certainty that you are at least as convinced by it as by any further evidence that you could find for it. Hence, evidence can no longer change anything about your belief. That is, once something has become certain, it is entirely removed from the push and pull of evidence. I take a functional notion of evidence here: Moot evidence – that is, evidence that does not actually change your confidence – is not evidence even though it counterfactually might be able to change your confidence. Such counterfactual evidence is certainly interesting and important, but in such counterfactual scenarios, I would have different certainties. Consequently, certainties are *actually* insensitive to evidence, even though there may be evidence that would *counterfactually* support them. Conditional certainties genetically depend on evidence insofar as they attained their certainty status through evidence – they are therefore supported by evidence in a certain sense.

Certainties therefore allow us to remain internally rational in the face of what is, from an external standpoint, adverse evidence. From your internal point of view, it is rational to dismiss evidence that, objectively speaking, contradicts a certainty. That is, certainties allow the subject to explain to themselves why they dismiss such counterevidence and they can also explain why such counterevidence is not even recognised as such.

You might wonder whether we ever are in such a state.<sup>12</sup> I will give some examples of such certainties and I will argue that they are an epistemically interesting phenomenon. In the literature too, certainty has recently been the focus of renewed attention as an epistemic state that deserves consideration, for instance, Beddor (2020).

### 1.3.1 *Anthropology of certainty*

Certainties are rife among even our everyday beliefs. We just rarely pay them any attention. Ludwig Wittgenstein's *On Certainty* (1969) presents a long list of things of which we are certain. I will cite some of the examples Wittgenstein gives of this phenomenon.

The absolute classic example for certainties is mathematical knowledge. You are dead certain that  $12 \times 12 = 144$  (Wittgenstein, 1969, §43). What sort of evidence could there be to show you that this is not the case? What other beliefs could you acquire that would bring you to doubt it? But also conversely: Could your confidence in your belief that  $12 \times 12 = 144$  be increased in any way? What would evidence confirming this belief look like? Your confidence in  $12 \times 12 = 144$  does not budge at least in ordinary contexts where you are not philosophising or drunk.<sup>13</sup>

Meanwhile, such unconditional psychological certainty is not present in more complex calculations, for example,  $236 \times 942 = 222,312$ . Here, we may only attain conditional certainty by diligently checking our calculations using different means. Yet we would become certain after checking it several times.

But according to Wittgenstein, certainty is not limited to mathematics. Another example is the belief that objects do not simply disappear and reappear. You are, for example, certain that you never disappeared into thin air while you slept only to reappear before waking up (Wittgenstein, 1969, §101). In the same vein, I am certain that the books I have stashed away in moving boxes do not disappear from time to time (Wittgenstein, 1969, §134). I expand on this example in the section on *baby logic* (3.4).

Again, no amount of evidence will shake our confidence in this certainty. Any apparently contrary evidence would be rejected as trickery or otherwise misleading. If I were to find that my books are no longer in their box, I would infer that someone took them. Even if I *saw* how someone sleeping disappeared into thin air, it would not change my certainty. I would believe that I just witnessed some contrived artifice, but the person still has to be somewhere and had to be moved there – stage magic is fed by the thrill of contradicting your certainties. Conversely, the fact that you have seen plenty of things without any of them ever disappearing arguably did not increase your confidence that they never do because you were already certain of this.

As a third and last example, consider our biological certainty that all animals have parents. Wittgenstein's example is that we are certain that cats do not grow on trees and that I have a mother and a father (Wittgenstein, 1969, §282). Arguably, you have witnessed very few reproductive processes and presumably this fact does not shake your confidence in the certainty that every animal has parents and ancestors. Conversely, you would reject any purported evidence as contrived or manipulated that seems to show that some animals lack ancestors or grow on trees.<sup>14</sup>

These examples show how diverse certainties may be and that they can be found in many different places. An interesting feature of certainties is that they are often only implicitly held. You may never have actually considered them as being propositions, but your confidence in them is unchanging (Wittgenstein, 1969, §159).

### 1.3.2 *Scepticism*

Another way in which certainties become apparent is when people are confronted with sceptical arguments. I want to argue that *psychologically* speaking many people accept the premises and inferences of sceptical arguments, but they do not follow through to their conclusion. Yet at the same

time they also do not abandon their premises. The sceptical argument presents them with a (doxastic) paradox (Wright, 1991, p. 89).

Consider Descartes' sceptical argument C: I cannot tell my current experience apart from a dream. If I cannot distinguish my experience from a dream, then I am not justified in my belief that I am not dreaming it. If I am not justified in my belief that my current experience is no dream, then I am not justified in accepting my perceptual beliefs based on this experience. Therefore, I am not justified in my perceptual beliefs (*This is my adaptation*; see Descartes, 2013, pp. 25–26).

At least some people will (psychologically) accept this argument's premises and inferences, but not the conclusion. They hold their perceptual beliefs as if they were justified. They are accordingly in a bind. This situation is made possible by the fact that sceptical arguments show that we lack justification or evidence for our certainties. But once something is a certainty, it does not require evidential support to be believed and it is not defeated by the absence of evidence. Thus, in the Cartesian example, I am simply certain that I can trust my experience.

To spell this out, take the example of some generic person S, some perceptual belief B, and the sceptical argument C:

- 1 S accepts the premises and inferential steps of sceptical argument C that shows that S's perceptual belief B lacks justification.
- 2 S infers that sceptical argument C shows that belief B lacks justification.
- 3 S is not moved by B's lack of justification and does not drop B.

This leads to the unhappy situation that S holds onto a belief, B, while also having drawn the conclusion that it lacks justification. This makes S appear very unreasonable, but if you grant that many people like S can swallow the sceptical argument without following through to its consequences, then we need to explain how this is possible.

The solution, I think is certainties:

- 4 S is *certain* that if she has perceptual experiences like that, then she is justified in accepting the corresponding belief B.<sup>15</sup>

Essentially, S takes it for granted that her experiences justify her beliefs. S begs the question against the sceptic, but not in the sense that S takes herself to *know* that she is justified in accepting B. Rather, S takes B to simply hold. S may say something like: 'Yes, the argument shows that I have no reasons to accept B, but I am certain that I can trust my experience. Therefore, I believe B.' I think this is the phenomenon that David Hume pointed to, namely how, outside the study, all those difficult sceptical problems become mere smoke and mirrors and do not play any role in daily life

(Hume, 2004, §126). Thomas Reid took a similar avenue (Reid, 2012, p. 171). This is not an epistemological response to the sceptical problem, but rather a psychological description of how people react to sceptical arguments. The fact that we are certain of things explains why scepticism is psychologically implausible (*cf.* Kekes, 1975). In Chapter 2, I will argue that such an agent also can be epistemically warranted in being certain that she is not victim of a sceptical scenario.

This account differs from the Moorean anti-sceptical strategy, although if you were to describe the behaviour of a Moorean and the ordinary person S, then you would hardly see a difference between them. The Moorean rejects the sceptical argument because they reject the argument's conclusion. A Moorean claims to *know* B, thus they are justified in accepting B. Therefore, they must reject the sceptical argument as either unsound or invalid (Hazlett, 2006, p. 211).<sup>16</sup>

The ordinary certainty that I describe is *not* a claim of knowledge. What S does by claiming certainty is something else: She ignores the problem. That is not an intellectually satisfying solution to the sceptical problem, but pragmatically it does the trick for S. I take a response to scepticism to be satisfying, if it does not beg the question against the sceptic or dismiss sceptical considerations out of hand (Stroud, 1984; Pritchard, 2002). In Chapter 5, I will propose such a solution. I believe this is a natural account of how many people react to a sceptical argument, given that they would not want to infer knowledge of the existence of an external world as the Moorean does.

### 1.3.3 *Deep disagreement*

If we go beyond a single individual's certainties, an interesting issue arises. It is not necessary that two persons share their certainties. So what happens if individuals S and T disagree about whether P or not P, while each is certain about their own claim?

Let us take the classic example of such a disagreement, religious matters. Consider two members of two small, related but fictional denominations: a very pious eastern Baptist and a very pious western Baptist. The two denominations are indeed very similar; they even have the same founder. However, the eastern Baptists use an early version of the founder's commentary on the Bible while the western Baptists use a later version.

Imagine these two meeting for breakfast without knowing about their differences. As they start their obligatory pre-breakfast ritual, they realise that there are subtle differences in their practice and they start arguing about the right way. Finally, they resort to citing the relevant religious texts of whose truth each is, respectively, certain. Only then do they realise that they are referring to different sources. However, each remains absolutely

committed to their version of the text. In the end, they decide to agree to disagree, each understanding that they will not be able to convince the other but remaining convinced that the other is a heretic.

More abstractly speaking, S and T would at first not realise that P is at issue, but start disagreeing about some other belief Q, which lies downstream from whether P or not, that is, some implications of P and some shared knowledge R. They would not even consider talking about P – if it is certain whether or not P, then it is also not worth mentioning. That is, being certain of something usually goes together with the assumption that the certainty is shared.

This disagreement then would lead to puzzlement as to how S and T can disagree about Q given that both agree about R. Only after a while would they come to realise that P is the object of their disagreement.

Given that each is certain of P, nothing that either can say will change the other's view. Given that they take P or not P to be indubitable, the fact that the other disagrees simply shows how mistaken the other is. Thus, their debate may devolve into a futile argument, each trying to raise points that are moot to the other.

That means they would reach an impasse where each recognises that their disagreement is irresolvable. They may at best agree to disagree, each recognising the other's certainty as a certainty they will not dislodge. By disagreeing about certainties, S and T have reached a natural end point.

We all have come across this fairly common phenomenon. It is called deep disagreement. Authors writing about the topic usually try to explain how agents can remain rational in the face of deep disagreement and they point to divergent certainties or hinges as an explanation (Fogelin, 1985; Hazlett, 2014; Ranalli, 2020). Lynch (2012) and Kappel (2012) show how deep disagreement can lead to far-reaching practical problems.

The issue here is that the reasonable default in the face of disagreement seems to be to lower one's confidence. This doctrine is called *conciliationism*. The idea here is that if you take the disagreeing other to know roughly as much about a matter as you, then the sheer fact that they can (and do) disagree with you gives you a reason to reconsider your belief because you might have overlooked something (Goldman and O'Connor, 2019). But if you are certain of your belief, then nothing will move you. Instead, you will infer from your certainty and the disagreement that the other is wrong – and consequently is not your epistemic peer. In sum, deep disagreement is one of the few cases where certainties are explicit and their dynamics can be studied, as they render conciliationism moot.

It can also be noted that deep disagreement is usually about fundamental beliefs or hinges. It is not about whether or not it rained yesterday: Everybody would agree that there is some way to settle that question. Rather, we usually deeply disagree about cornerstones or rules that have

wide-reaching implications. There are then pedestrian disagreements that can be settled by evidence and deep disagreements about hinges which cannot (*cf.* Coliva and Doulas, 2022).

#### 1.3.4 *Baby logic*

In this section, I argue that there are some certainties that we are born with. More exactly, they are preprogrammed. I refer here to *core cognition* or *core knowledge* (Spelke and Kinzler, 2007; Carey, 2009). Core knowledge consists of innate conceptual resources that organisms rely on without having learned them – they are hardwired and invariable. One way of interpreting these features is as certainties.

Core knowledge can be described as consisting of beliefs that we find ourselves having due to our biological make-up. Susan Carey argues that it is based on the so-called dedicated input analysers (DIA). These interpret a narrow range of sensory inputs and produce corresponding classifications (Carey, 2009, pp. 10–11). One important function of DIAs is to give (non-adult) organisms cognitive guard rails to cope with their environment: Consider how ducklings imprint on their mother when born. Human infants are strongly predisposed to pay attention to faces, and the fundamental mechanical beliefs we are born with also belong to our core cognition, for example, that solids cannot pass through each other.

Why should these innate cognitive features be termed certainties? The key point is that they are considered as being evidentially insulated. Their status is not influenced by other things we believe. Carey describes core cognition as ‘modular and informationally encapsulated, protected from explicitly held conceptual knowledge’ (Carey, 2009, p. 461). This means that whatever I believe will not change my core beliefs. This is also the feature that I argued characterises certainties. Whatever else we learn, a piece of our core knowledge will stubbornly remain operative – it is hardwired.

You might object that core cognition cannot consist of certainties because it does not consist of beliefs. DIAs are not beliefs, while certainties are beliefs. Therefore, core knowledge does not consist of certainties.

First, I want to point out that the term ‘core knowledge’ (Spelke and Kinzler, 2007) actually points towards core cognition as consisting of beliefs and even certainties. However, this is only circumstantial evidence and hardly compelling regarding the status of core cognition as certainties.

Second, I also want to respond that DIAs behave much more like beliefs than it would seem at first sight. As a point in case, core cognition is inferentially integrated:

However, representations in core cognition differ from sensory and perceptual representations in having a rich, conceptual, inferential role to

play in thought, even infants' thought. Representations that are the outputs of distinct core cognition systems are inferentially integrated and are in this sense central.

(Carey, 2009, p. 11)

Admittedly, DIAs are not inferentially integrated, only their outputs. But they do produce beliefs. I want to argue that given their structure, we can nevertheless also consider DIAs as beliefs. This stretches the notion of belief a bit, given that input analysers are usually conceived of as functions with a sensory input that have representational states as their output. However, in this line, Zoe Jenkin (2020) argues that our core cognitive constraints – that is, our core knowledge – serve as justifying reasons for our core cognitive output representations.

We may minimally understand by 'commitments' or 'beliefs' representational states that present their content as true and thereby influence behaviour and the formation of further beliefs *as well as* their functional equivalents. DIAs clearly do not correspond to some mental item that could be found through introspection and that is a certain propositional belief or acceptance. However, given their structure, we can conceive of their having a correlate belief with the fairly complex content that *<this sort of sensory input corresponds to that sort of external object/behaviour>*. Calling these structures beliefs generates a deflationary notion of belief (Carey, 2009, p. 96).

To account for this, I would like to introduce a distinction between explicitly believed certainties that can be verbalised and implicit certainties. Implicit certainties are acceptances that we need not be aware of holding, but that stand firm given the other things we believe and the sorts of things we infer – our implicit premises. Implicit certainties are the invisible but necessary background to our epistemic activity.

This implicitness is a key feature of some debates on certainties. With reference to the implicit certainty 'I have a body,' Danièle Moyal-Sharrock talks of

a disposition of a living creature which manifests itself in her *acting in the certainty of having a body* . . . for example, in her eating, running and her not attempting to walk through walls as if she were a disembodied ghost.

(Moyal-Sharrock, 2004, p. 67)

This notion of certainties clearly fits very well with Carey's account. For example, it is part of our core cognition that objects do not pass through each other (Carey, 2009, pp. 41–42). This idea also accords with Annalisa Coliva's thesis that some of our fundamental certainties are *rules* and these



are therefore not regularly held beliefs (Coliva, 2015, p. 10). We may follow rules without being able to verbalise them. Coliva argues that rules aren't even truth-apt. Conceiving certainties as rules seems to fit with the notion of a DIA. I will come back to Coliva's and Moyal-Sharrock's views and DIAs in the section on hinge certainties. I disagree with them about whether these propositions are believed and truth-apt.

Note too that the inferential integration of core cognition means that core knowledge does influence behaviour and other beliefs by producing regular beliefs (Jenkin, 2020). The point where core knowledge differs from regular beliefs in its interaction with other beliefs lies with its not being influenced by evidence or other beliefs – just like certainties.

I argue that if we grant infants beliefs in a deflationary sense, then dedicated input analysers are unconditional certainties. They just are there: Neither animals nor neonates base them on anything else. We find ourselves saddled with core cognition. According to Carey, we still possess core cognition modules even as adults. We perceive causality like infants do (Carey, 2009, p. 461) and even if we know better, at some level we fall victim to the Muller-Lyer illusion. However, this primitive layer is overlaid with further layers of more sophisticated physical and geometrical knowledge.

Hinge epistemologists have, until now, mostly bracketed the question of whether there are innate or biologically anchored certainties. Wittgenstein (1969) himself emphasises the social aspects of the acquisition of hinges, using often examples of children acquiring hinges. Also, Moyal-Sharrock focuses more on social aspects but raises the possibility that some certainties are anchored in an innate trust (Moyal-Sharrock, 2004, p. 200). To my knowledge, there has not been any explicit argument that there are no innate certainties. I would argue that this is an interesting expansion of hinge epistemology especially because there are informative empirical studies on how beliefs or certainties manifest during development.

John Greco also has recently made this cognitive turn in hinge epistemology by arguing that our hinges consist in *procedural knowledge*. Procedural knowledge is the 'knowledge exercised in the performance of some task' (Greco, 2021, p. 114). On this conception, core cognition definitely is procedural knowledge, but not *vice versa*. Procedural knowledge need not be biologically anchored, although Greco (2021, p. 118) admits of this possibility. Like me, Greco (2021, p. 121) points out that procedural knowledge is certain in the sense of being evidentially insensitive or hardwired. Our major and important difference is that Greco argues that these certainties are indeed *knowledge*, manifesting epistemic competence (Greco, 2021, p. 123),<sup>17</sup> while I follow Wittgenstein and Pritchard in the view that we do not *know* our hinges.

Thus, if you grant that core cognition consists of beliefs and include the observation that these beliefs are evidentially insulated representations, then it follows from this that core cognition consists of certainties. Notably, some certainties are innate, that is, are not learned and these help us to come to cognitive grips with the world. Additionally, some of these innate certainties appear to accompany us all our lives.

#### 1.4 Hinge certainties

I now want to turn my attention from how hinge propositions relate to other accepted propositions to how they function as the content of beliefs or acceptances. I argued that hinge propositions are necessarily accepted or believed non-evidentially because any evidence in favour of a hinge would necessarily be circular.

I will now further argue that, given their peculiar position, hinge propositions need to be accepted with certainty. That is, in a belief set, hinge propositions have to be insensitive to evidential defeat (and support). I therefore call our beliefs in hinges *hinge certainties*, because we will always be certain of hinges.

##### 1.4.1 *Why we are certain of our hinges*

If hinge beliefs were not certainties, that is, if they could be defeated by evidence, then they would not actually be central cornerstone propositions, and they would not function as rule propositions that express the rules by which we update our belief set or carry out our actions, because they then would only appear to be the central presuppositions for our set of accepted propositions. In the complete absence of hinge certainties, our belief systems would be in permanent upheaval since we would lack stable rules for changing our belief set and our beliefs would continually shift in their content because there would be no cornerstones. Such an unstable belief set would be pathological, a symptom of mental illness (Gipps and Rhodes, 2008).

The implicational structure of a belief system is the first reason for thinking that cornerstones and rules are certainties. It is in the nature of the closure of entailment that the consequent requires at least as strong doxastic confidence as the antecedent.<sup>18</sup> This also counts for the degree of certainty, that is, the extent to which a conviction is sensitive to evidence. Granting that there are (also non-hinge) certainties, as was argued earlier, any belief entailed by this certainty, notably any more central belief, will also be certain. Consequently, all cornerstones entailed by certainties will be insensitive to evidence. Note that this does not imply that all certainties are hinges (Smith, 2019).<sup>19</sup>

The fact that certainties and thereby hinge certainties are insensitive to evidence also fits well with the fact that hinge propositions determine evidential relations. Rule propositions determine what counts as evidence. Hinge propositions would be self-defeating if they allowed for evidence that rebuts or undermines them because, as hinges, they determine what counts as evidence for or against propositions, including themselves. Consequently, either these are not actually hinges or they are certainties that cannot be defeated by any counterevidence. Hence, the evidence-determining role of rule propositions implies that we are certain of hinge propositions. For this reason, I call their acceptance hinge certainties.

This means that we now have arguments for why some certainties are hinges as well as an argument for why all hinges are certainties. I take it that Wittgenstein gestured at this sort of consideration when he noted that giving evidence must come to an end at some point (Wittgenstein, 1969, §204). Some things are so fundamental that we must be certain of them, thus removing the need for further reasons. This also explains why track record arguments for our hinges as discussed earlier are rendered moot even if they worked: They rely on your certainties to raise your confidence in your certainties which cannot be raised. The argument does no argumentative work.

In the section on certainties, I argued that these allow us to avoid sceptical conclusions. This view can now be supplemented with the observation that sceptical arguments usually target hinge propositions (Davies, 2004, p. 213). That is how these arguments have their reach and can seek to undermine large swaths of knowledge. We can sidestep the sceptical conclusion by being certain of the undermined proposition, that is, being certain about hinges. This thus provides further corroborating evidence for the claim that hinges are indeed certainties.

Wittgenstein also argued that hinges are the answer to sceptical arguments. His own argument was based on targeting the epistemic state of doubting. We can take the sceptical argument as advocating doubting one of our hinges and all the propositions dependent on them. *Qua* epistemic state that relates to other epistemic states, doubt too relies on some hinges that stay put, that is, are certain. Were this not the case, then the act of doubting would be self-undermining and not actually an act of doubt (Wittgenstein, 1969, §24). Thus the very act of doubting commits you to accepting some hinges that are necessary preconditions for being able to doubt.

I also would like to return to consideration of *core cognition* as a set of innate certainties. Core cognition is the hardwired set of certainties that constitutes an infant's starter kit in getting to know the world. That is, core cognition is the set of central beliefs, the cornerstones of the infant's knowledge of the world. Without an innate core cognition, the infant would be helpless when it comes to interpreting the empirical information she receives.

The motivation behind theories of core cognition is that we need a sort of cognitive toolbox to begin grasping our environment. These theories are opposed to *blank slate* theories, which assume that neonates possess no conceptual resources at all. If neonates really did have no beliefs at all, they would lack any resources for interpreting their environment. They would just experience an unstructured stream of impressions. Innate certainties determine the content of and the rules for the beliefs we form, accordingly they are not mere certainties but *hinge certainties*. For example, neonates would be unable to cognise objects (Carey, 2009, p. 27) without already possessing a concept, that is, without possessing the hinge that ‘my environment consists of concrete continuously existing objects.’

Core cognition gives neonates the hinge certainties that they need to come to grips with the world. It is through them that they know how to pay attention to faces and it gives them the cues needed in recognising their primary carer. Further, if we grant that perception is cognitively penetrable (Stokes, 2013), then our hinges as encoded in core cognition will also influence how we perceive our environment (*cf.* Jenkin, 2020).

Essentially, core cognition gives the foundation on which the neonates’ other beliefs are built. Core cognition therefore has much the same function as hinge propositions. I also argued earlier that it is plausible to interpret core cognition as consisting of certainties, which would fit with its hinge function. I take Danièle Moyal-Sharrock’s (2004) theory of animal hinges to be an account of core cognition.

#### 1.4.2 *Two simple examples*

The belief that the Palace of Westminster stands on the Thames presupposes the fact that objects can stand in spatial relations. That is, by accepting the former, you need to accept the latter. Consequently, one of your cornerstones is that there is such a thing as spatial relations, you are certain of this. To see that this is a cornerstone proposition, consider the sort of reason you could give for the existence of spatial relations: All instances of ‘being on,’ ‘being next to,’ or ‘being behind,’ and so on presuppose that things are spatially related.

You might now argue that you experience things as spatially related. This again presupposes that your experience of things is indeed accurate and correctly conceptualised. This is a rule proposition that tells you how to implement your experience. Again you believe that your spatial experience tracks how things are, among others, experience tracks that there are indeed spatial relations.

Now consider how deeply this runs: Could you grasp what it might mean if someone told you that, while both exist, the Thames and Westminster Palace do not stand in any spatial relation to each other? I, personally,

am not sure what the person could even mean by such a statement. I would indeed reject any evidence that they would proffer. My belief that any two extant objects stand in a spatial relation stands firm for me – I am certain of the hinge proposition that things stand in spatial relations.

To make an analogy: On a naïve level, we seem to experience the stars as being all at a roughly equal distance from us and distributed across a dome. Someone might have a hinge that this experience is accurate and they would do fine in everyday life. They might also be utterly unable to conceive of the extra-lunar sphere as extended space – too much in their world view would have to be remodelled for that to make sense. They are therefore certain of the hinge that we live under a celestial dome.

This small example of space and how we experience it shows how tightly interwoven the certainties are that flow from content and those we need as a basis to change our beliefs. What we take to be the case content-wise influences how we experience things and *vice versa*. Interestingly, there is not one universal human set of such certainties. There may be a minimal set shared by everyone, for instance, that our experience tracks the world's state, that there are other people, and so on, but this minimal set may be smaller than we expect. I will deal with this question in the third chapter.

## 1.5 Trust

As a last step, I will argue that hinge certainties, or certainties about hinge propositions, form a distinct class of doxastic states. Namely, if I am certain of a hinge, then I *trust* it to be the case (Wright, 2004).

This does not mean that all cases of trust are certainties about hinges, not by far. Rather it means that certainties about hinges share essential features with other kinds of trust. I shall call this kind of trust, that is, certainties about cornerstones and rules, *epistemic trust*. Arguably, it would be better described by designating it fundamental epistemic trust,<sup>20</sup> but I will also argue that epistemic trust always needs to be fundamental, that is, to be about central propositions (Moyal-Sharrock, 2004, pp. 193–195); otherwise, it is merely pragmatic.

This notion of trust is fairly close to Cohen's (1989) notion of acceptance. Notably, I take trusting to be a frequently active affair. I may (decide to) trust that something is the case even though I feel in no way compelled to believe it. My notion of trust also fits with Cohen's remark that acceptances are stronger than mere suppositions (*cf.* Engel, 2012). Meanwhile, there is also implicit trust.

Why would trust in cornerstones and rules be epistemic? I would argue that these certainties essentially aim at how the world is. That is, hinge certainties aim at truth. Their role is to model the world, nothing else. This truth-directedness makes trust in hinges epistemic.

Given this truth-directedness, trust in hinge propositions is more than a mere pragmatical reliance on the proposition in order to achieve some further purpose. Meanwhile, Wright (2014, p. 226) argues that we need to trust hinges because unconditionally believing them to be true would be irrational.

An obvious *caveat* here is that this notion of trust is not a trust in persons, but rather a trust in propositions. It is a doxastic state, not an attitude towards people. I call these two kinds of trust doxastic and social trust, respectively. However, as I will argue, there is a corresponding doxastic state for any trusting attitude towards a person.

Why should my *certainty* about a cornerstone be a case of trust? I am certain, why would I need to *trust*? The notion of trust implicit in this question assumes that trust is accompanied by a fairly low degree of confidence. When trusting, I am aware that it is *mere trust*, there is no evidence or anything else supporting it; thus, I have a lesser degree of confidence in what I trust in. This reasoning, however, does not describe actual trust, but rather *mistrust* or at least the absence of trust.<sup>21</sup> Moyal-Sharrock (2004, p. 194) goes even further: As soon as we start thinking about the object of our trust, we stop trusting it.

Trust may indeed be accompanied by very high, even absolute degrees of confidence, certainty in other words. This sense can be found in the notion of ‘blind trust.’ *Blind* trust obviously also seems to imply that something has gone epistemically wrong for you, because nothing can change your mind about that which you trust in.

One of the key aims of this work is to argue that there is a kind of trust with absolute confidence, that is, blind trust, that is epistemically appropriate. There is a way to *trust responsibly*, without incurring the epistemic defect apparently implied by its blindness.

Trust as a doxastic state is characterised by its not requiring any evidence to be adopted. I trust a proposition P to be the case if and only if I have no sufficient evidence for P but I accept P nevertheless as true (Wittgenstein, 1969, §672). Cohen’s (1989) notion of acceptance also allows for adoption independently of evidence, but it does not require it.

Hinge certainties do not require sufficient evidence either. What distinguishes them from paradigmatic cases of trust is that I may not necessarily be aware that I lack sufficient evidence for a certainty. However, I would argue that the same may hold in cases of trust. People may trust something to be the case without actually being aware that they are merely trusting it.

This means that non-evidential certainty is a subclass of the class of trust. There may be certainties that are based on evidence. But as I argued earlier, certainty concerning cornerstones and rules is not and cannot be supported by non-circular evidence. Hence, accepting hinges is clearly an instance of trust.

Danièle Moyal-Sharrock (2004) defends an even stronger version of epistemic trust in hinges, *ur-trust*. According to her, trust must be necessarily blind because it is non-cognitive, non-propositional, indubitable, grammatical, and foundational (Moyal-Sharrock, 2004, pp. 195–196). Differently from my conception of epistemic trust, Moyal-Sharrock’s *ur-trust* is completely subtracted from the push and pull of our doxastic life. As soon as we even consider its objects, it is destroyed (Moyal-Sharrock, 2004, p. 194). I would argue that our trust in hinges survives our thinking about them and how they relate to each other.

Moyal-Sharrock also examines what happens when we stop trusting our hinges. This lack of trust is *pathological*. When you cannot trust a hinge, then this leads to mental illness (Moyal-Sharrock, 2004, p. 197). I find it highly plausible that illnesses like schizophrenia are accompanied by a loss of trust in our hinges.

I also wrote that the only kind of epistemic trust is in fundamental hinge propositions. This means that I deny that trust in any peripheral proposition can be epistemic. If we accept peripheral propositions without evidence, it is because we rely on them for pragmatic purposes. Trust in peripheral propositions is always based on pragmatic, not epistemic, reasons.

Why is it only trust in hinges that is epistemic? As mentioned before, I cannot help but trust in cornerstones and rules. There is no other way of acquiring non-circular evidence for them therefore the only thing I can do is trust that they are true.

This does not hold for peripheral propositions. They are characterised by the fact that they can be supported or challenged by evidence. There is in principle sufficient evidence to be had for some peripheral proposition. I can be *certain* of that proposition because the available evidence was sufficient, definitive, and overwhelming (Beddor, 2020). But that does not make the certainty into a hinge certainty or into a case of trust, because it still needs evidence and it simply happens to be the case that the evidence is so great that the proposition is taken as certain.

But given that there is in principle always evidence available for peripheral propositions, evidence becomes the epistemically optimal way to settle questions concerning them. Whether or not peripheral propositions are true is essentially an empirical<sup>22</sup> question. We settle empirical questions by evidence and sufficient non-circular evidential support is the gold standard in internalist epistemology.

Hence, not basing your peripheral beliefs on evidence is epistemically suboptimal. That is, trusting something to be the case when you could in principle find out whether or not it is the case is not epistemically optimal. This therefore cannot be *epistemic* trust.

That does not mean that it is unreasonable to trust in general or rely on peripheral propositions. You may lack resources such as the time or cognitive capacity to figure out whether something is the case or not<sup>23</sup> while still needing to decide about it to avoid getting epistemically or practically stuck.

These are cases like the road to Larisa (*cf.* Plato, 2005), where you have to pick which road leads to Larisa. You have no evidence whatsoever, but in theory there might be evidence available. You nonetheless need to trust in and rely on the belief that one of the two roads leads to Larisa, because you need to get to Larisa.

This form of trust clearly is not epistemic, but purely pragmatic. Any forthcoming evidence should be capable of changing your mind about whether the road you picked leads to Larisa. For this reason, trust in or reliance on peripheral propositions can only be pragmatic, that is, subject to some practical interest. In order for it to be epistemic, there must be no prospect of getting any non-circular evidence.

Finally, social trust, namely trust towards people, is always accompanied by an instance of trust as a doxastic state. This does not mean that social trust is always epistemic trust, it may be simply pragmatic. It instead means that if I trust you to do something, that is, if I have a person-directed trust attitude, then I also trust in the truth of the correspondent proposition that you will do it. That is, I will also have a corresponding attitude of doxastic trust.

While social trust in a person always implies doxastic trust that something is the case, the converse is not necessarily true. More specifically, it is not the case that every instance of trust in propositions involving persons will be an instance of trust in these persons. I may, for example, trust that Donald Trump will bungle some project without having the least amount of trust in Donald Trump.<sup>24</sup>

## 1.6 Trust in hinge certainties

In this chapter, I introduced the concepts of hinge, certainty, and trust. I argued that these are built up on each other. First, I introduced the concept of a hinge, that is, propositions that fulfil an important enabling function within a belief system. Second, I introduced the concept of doxastic certainty, beliefs that cannot be defeated by contrary evidence. I illustrated this concept with a wide range of examples. I then argued that hinges need to be held with certainty in order to fulfil their function in a stable way. The fact that hinges are certainties also explains how they relate to the problem of scepticism. Finally, I argued that hinge certainties are best conceived of as instances of epistemic trust. That is, the technical concept of a hinge certainty coincides with the concept of trust that something is the case.



Table 1.1 Types of hinge acceptances

	<i>Belief</i>	<i>Certainty</i>	<i>Trust</i>
<b>Regular proposition</b>	Ordinary belief	Conditional certainty	Pragmatic trust
<b>Hinge proposition</b>	Trust in unconditional hinge certainties		

## Notes

1. I mean by this that there could be two doxastically identical agents who are nevertheless in very different epistemic situations. Both believe the same things for the same internal reasons but one may get things very wrong because their epistemic environment is hostile while the other is broadly right and possesses considerable knowledge. This phenomenon underlies the new evil demon problem. See, for example, Foley's (1993) notion of 'egocentric rationality.'
2. This is a strong notion of certainty and I will argue further for it in Section 1.3. Note that we may lose certainties through non-epistemic processes, that is, psychological non-cognitive processes like forgetting or through emotional responses.
3. Wishful thinking is an irrational kind of trust: it requires specific desires that something be the case, which trust in general does not.
4. Wright does not take a strong position on his theory of belief and acceptance. For some detailed accounts of acceptance, see Lehrer (1983), Cohen (1989), and Bratman (1992).
5. Carroll problematises the other direction of implication: Just because I accept a proposition formulating the rule, I am not committed to following the rule.
6. An individual who has been isolated from birth will nevertheless have beliefs, but they will be incomparably more primitive than our shared frameworks.
7. This type of two-level account will play a *different* role later in the book.
8. These do not map onto the four alternatives to scepticism from Van Cleve (2003).
9. This criterion serves to exclude arational mechanisms like a shock or brain-washing. Arational mechanisms are probably more prevalent than we would like to think – advertisement and propaganda exploit this, for example.
10. There is also the content sense of 'certainty,' that is certainties can be interpreted as the propositions that we believe with certainty. For clarity's sake, I will avoid using 'certainty' in this content sense.
11. At least on standard accounts.
12. Note that proving that some supposed certainty is defeated by another belief does not yet show that there are no certainties. You have to be more certain about the defeater than about the defeated belief – in such a case, you were simply mistaken about what you are certain of.
13. Note that just because  $12 \times 12 = 144$  is a certainty, does not mean that it also is about a hinge. I consider the role that certainties play in philosophy in Chapter 2.
14. Though notably, for a long time, the spontaneous creation of life was thought to be common. Note also that both artificial and natural clones (through parthenogenesis) have a single parent.
15. There may be other ways in which S could claim justification. However, usually this requires a fair degree of philosophical sophistication and I do not

think everyone will be able to avail themselves of such strategies. The only thing I want to show here is that it is fairly commonplace to ignore the apparent lack of justification for a claim due to certainty.

16. I thank an anonymous referee for pointing out that there is a version of the sceptical argument that challenges the rationality of S's beliefs which S would reject because she is certain. This fits well with Coliva's (2015) and Hazlett's (2006) positions – which I will consider in Chapter 2. I want to note that appreciating the difference between this sceptical argument and other versions already requires quite some philosophical sophistication which I assume to be missing in this case.
17. We will return to this in Chapter 5.
18. If I were less confident in belief B's recognised necessary presupposition H than in B, then I should, rationally speaking, adjust my confidence in either H or B. *De facto*, I think many people simply possess many certainties.
19. Thanks to Nikolaj Pedersen for raising this point.
20. Compare this to the German notion of '*Urvertrauen*.'
21. Compare this also to William James' (2014) suggestion that we trust in God and his existence.
22. I am stretching the notion of the empirical to some extent here by seeing all evidence as falling under the notion of experience or *empirya*.
23. With hinges, no amount of cognitive capacity or time would furnish you with evidence.
24. Annalisa Coliva makes a similar argument that social trust can be explained in terms of *hinge trust*, showing how this concept can be found in Wittgenstein's *On Certainty* (Wittgenstein, 1969). She argues that trust is a fundamental type in our epistemology and that it cannot be analysed in terms of other concepts (*private communication*).

## References

- Anscombe, G. E. M. (1957) *Intention*. Cambridge, MA: Harvard University Press.
- Beddor, B. (2020) 'New Work for Certainty', *Philosophers' Imprint*, 20, pp. 1–25.
- Bergmann, M. (2004) 'Epistemic Circularity: Malignant and Benign', *Philosophy and Phenomenological Research*, 69(3), pp. 709–727. <https://doi.org/10.1111/j.1933-1592.2004.tb00524.x>.
- Block, N. (1998) 'Semantics, Conceptual Role', in *Routledge Encyclopedia of Philosophy*. London: Routledge. <https://doi.org/10.4324/9780415249126-W037-1>.
- Braithwaite, R. B. (1968) *Scientific Explanation: A Study of the Function of Theory, Probability and Law in Science*. Cambridge: Cambridge University Press.
- Bratman, M. E. (1992) 'Practical Reasoning and Acceptance in a Context', *Mind*, 101(401), pp. 1–16.
- Brown, C. (2022) 'Narrow Mental Content', in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* (Summer 2022). Available at: <https://plato.stanford.edu/archives/sum2022/entries/content-narrow/>.
- Carey, S. (2009) *The Origin of Concepts*. Oxford: Oxford University Press.
- Carroll, L. (1895) 'What the Tortoise Said to Achilles', *Mind*, 4(14), pp. 278–280.
- Chalmers, D. J. (2003) 'The Nature of Narrow Content', *Philosophical Issues*, 13(1), pp. 46–66. <https://doi.org/10.1111/1533-6077.00004>.
- Cohen, L. J. (1989) 'Belief and Acceptance', *Mind*, 98(391), pp. 367–389.

- Coliva, A. (2015) *Extended Rationality*. Basingstoke: Palgrave Macmillan.
- Coliva, A. and Doulas, L. (2022) ‘What Philosophical Disagreement and Philosophical Skepticism Hinge on’, *Synthese*, 200(3), p. 251. <https://doi.org/10.1007/s11229-022-03735-6>.
- Coliva, A. and Moyal-Sharrock, D. (eds.) (2016) *Hinge Epistemology*. Leiden: Brill.
- Davies, M. (2004) ‘II – Martin Davies: Epistemic Entitlement, Warrant Transmission and Easy Knowledge’, *Aristotelian Society Supplementary Volume*, 78(1), pp. 213–245. <https://doi.org/10.1111/j.0309-7013.2004.00122.x>.
- Descartes, R. (2013) *Meditations on First Philosophy*. Edited by J. Cottingham. Cambridge: Cambridge University Press.
- Dretske, F. (2005) ‘Is Knowledge Closed Under Known Entailment?’, *Contemporary Debates in Epistemology*, pp. 13–26.
- Engel, P. (2012) ‘Trust and the Doxastic Family’, *Philosophical Studies*, 161(1), pp. 17–26. <https://doi.org/10.1007/s11098-012-9946-z>.
- Fogelin, R. (1985) ‘The Logic of Deep Disagreements’, *Informal Logic*, 7(1), pp. 1–8. <https://doi.org/10.22329/il.v7i1.2696>.
- Foley, R. (1993) *Working Without a Net*. Oxford: Oxford University Press.
- Gipps, R. G. T. and Rhodes, J. (2008) ‘The Background Theory of Delusion and Existential Phenomenology’, *Philosophy, Psychiatry, & Psychology*, 15(4), pp. 321–326. <https://doi.org/10.1353/ppp.0.0208>.
- Goldberg, S. (2015) ‘Epistemic Entitlement and Luck’, *Philosophy and Phenomenological Research*, 91(2), pp. 273–302. <https://doi.org/10.1111/phpr.12083>.
- Goldman, A. I. and O’Connor, C. (2019) ‘Social Epistemology’, in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2019). Available at: <https://plato.stanford.edu/archives/fall2019/entries/epistemology-social/>.
- Greco, J. (2021) *The Transmission of Knowledge*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108560818>.
- Hazlett, A. (2006) ‘How to Defeat Belief in the External World’, *Pacific Philosophical Quarterly*, 87(2), pp. 198–212. <https://doi.org/10.1111/j.1468-0114.2006.00255.x>.
- Hazlett, A. (2014) ‘Entitlement and Mutually Recognized Reasonable Disagreement’, *Episteme*, 11(1), pp. 1–25. <https://doi.org/10.1017/epi.2013.46>.
- Hume, D. (2004) *An Enquiry Concerning Human Understanding*. Mineola: Dover Philosophical Classics.
- James, W. (2014) ‘The Will to Believe’, in *The Will to Believe: And Other Essays in Popular Philosophy*. Cambridge: Cambridge University Press, pp. 1–31.
- Jenkin, Z. (2020) ‘The Epistemic Role of Core Cognition’, *Philosophical Review*, 129(2), pp. 251–298. <https://doi.org/10.1215/00318108-8012850>.
- Kappel, K. (2012) ‘The Problem of Deep Disagreement’, *Discipline Filosofiche*, 22(2), pp. 7–25. <https://doi.org/10.1400/205566>.
- Kekes, J. (1975) ‘The Case for Scepticism’, *The Philosophical Quarterly*, 25(98), pp. 28–39. <https://doi.org/10.2307/2217950>.
- Lehrer, K. (1983) ‘Belief, Acceptance and Cognition’, in Parret, H. (ed.) *On Believing. Epistemological and Semiotic Approaches*. Berlin: Walter De Gruyter, pp. 172–183. <https://doi.org/10.1515/9783111727967-012>.
- Lopez Barranco, A. (2023) ‘Hinge Commitments as Arational Beliefs’, *Synthese*, 201(109). <https://doi.org/10.1007/s11229-023-04090-w>.

- Lynch, M. P. (2012) *In Praise of Reason*. Cambridge, MA: MIT Press.
- Lynch, M. P. and Silva, P. (2016) 'Why Worry About Epistemic Circularity?', *Journal of Philosophical Research*, 41, pp. 33–52. <https://doi.org/10.5840/jpr20165362>.
- Moyal-Sharrock, D. (2004) *Understanding Wittgenstein's on Certainty*. Basingstoke: Palgrave Macmillan.
- Ohlhorst, J. (2021) 'The Certainties of Delusion', in Moretti, L. and Pedersen, N. J. L. L. (eds.) *Non-Evidentialist Epistemology*. Leiden: BRILL, pp. 211–229. [https://doi.org/10.1163/9789004465534\\_012](https://doi.org/10.1163/9789004465534_012).
- Plato (2005) *Plato: Meno and Other Dialogues*. Edited by R. Waterfield. Oxford: Oxford University Press.
- Pritchard, D. (2002) 'Resurrecting the Moorean Response to the Sceptic', *International Journal of Philosophical Studies*, 10(3), pp. 283–307. <https://doi.org/10.1080/09672550210152122>.
- Pritchard, D. (2016) *Epistemic Angst*. Princeton: Princeton University Press.
- Quine, W. V. O. and Ullian, J. S. (1978) *The Web of Belief*. New York City: McGraw-Hill. <https://doi.org/10.1017/CBO9781107415324.004>.
- Ranalli, C. (2020) 'Deep Disagreement and Hinge Epistemology', *Synthese*, 197(11), pp. 4975–5007. <https://doi.org/10.1007/s11229-018-01956-2>.
- Reid, T. (2012) *An Inquiry Into the Human Mind*. Cambridge: Cambridge University Press.
- Schmitt, F. F. (2004) 'What Is Wrong With Epistemic Circularity?', *Philosophical Issues*, 14(1), pp. 379–402. <https://doi.org/10.1111/j.1533-6077.2004.00035.x>.
- Schwitzgebel, E. (2019) 'Belief', in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2019). Available at: <https://plato.stanford.edu/entries/belief/>.
- Smith, M. (2019) 'Full Blooded Entitlement', in Graham, P. J. and Pedersen, N. J. L. L. (eds.) *Epistemic Entitlement*. Oxford: Oxford University Press, pp. 281–295. <https://doi.org/10.1093/oso/9780198713524.003.0010>.
- Spelke, E. S. and Kinzler, K. D. (2007) 'Core Knowledge', *Developmental Science*, 10(1), pp. 89–96. <https://doi.org/10.1111/j.1467-7687.2007.00569.x>.
- Stokes, D. (2013) 'Cognitive Penetrability of Perception', *Philosophy Compass*, 8(7), pp. 646–663. <https://doi.org/10.1111/phc3.12043>.
- Stroud, B. (1984) 'Skepticism and the Possibility of Knowledge', *The Journal of Philosophy*, 81(10), pp. 545–551.
- van Cleve, J. (2003) 'Is Knowledge Easy – or Impossible? Externalism as the Only Alternative to Skepticism', in Luper, S. (ed.) *The Sceptics*. Aldershot: Ashgate, pp. 45–59.
- van Cleve, J. (2013) 'Why Coherence Is Not Enough: A Defense of Moderate Foundationalism', in Steup, M., Sosa, E., and Turri, J. (eds.) *Contemporary Debates in Epistemology*. Oxford: Blackwell, pp. 255–267.
- Vogel, J. (2000) 'Reliabilism Leveled', *The Journal of Philosophy*, 97(11), pp. 602–623. <https://doi.org/10.2307/2678454>.
- Wittgenstein, L. (1958) *Philosophical Investigations*. Oxford: Blackwell.
- Wittgenstein, L. (1969) *Über Gewissheit – On Certainty*. Edited by G. E. M. Anscombe and G. H. von Wright. Oxford: Blackwell.
- Wright, C. (1991) 'Imploding the Demon', *Mind*, 100(1), pp. 87–116.
- Wright, C. (2004) 'I – Crispin Wright: Warrant for Nothing (and Foundations for Free)?', *Aristotelian Society Supplementary Volume*, 78(1), pp. 167–212. <https://doi.org/10.1111/j.0309-7013.2004.00121.x>.

- Wright, C. (2014) 'On Epistemic Entitlement (II)', in Dodd, D. and Zardini, E. (eds.) *Scepticism and Perceptual Justification*. Oxford: Oxford University Press, pp. 213–247. <https://doi.org/10.1093/acprof:oso/9780199658343.003.0011>.
- Yli-Vakkuri, J. and Hawthorne, J. (2018) *Narrow Content*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198785965.001.0001>.

## 2 Entitlement

### 2.1 Introduction

In the preceding chapter, I introduced a class of doxastic states: Trust in hinge certainties. I argued that we accept hinge propositions as certain even though they are not analytic statements and cannot be supported by non-circular evidence. In this chapter, I will examine why and how this trust in hinges can be epistemically warranted. I will argue that we are warranted to trust in hinges by being epistemically *entitled* to do so (Wright, 2004).

My approach to epistemic entitlement, beginning with trust as a doxastic state and examining its epistemic warrant, is different from how things are usually argued for. Historically, the dialectic of the entitlement approach is to furnish an internalist response to the sceptical problem: We are warranted to trust in hinges independently of the evidence that the sceptic claims we lack (Wright, 2004). Thus, normally, the sceptical problem is introduced and then entitlement is presented as the internalist cure. While the sceptical problem plays an important role in my account, I also examine the broader role of epistemic trust. In other words, I think that epistemic entitlement is more than just an anti-sceptical strategy; it also gives us a framework for epistemically evaluating usually overlooked doxastic states, namely hinge certainties. Not all hinge certainties show up in sceptical arguments.

I argued that fundamental epistemic trust plays an important role in our doxastic households. It lurks in the background of all our regular beliefs. If we were not disposed to be certain of these propositions, our other beliefs would crumble. Additionally, I have argued in a naturalist vein that given our cognitive psychology, we are in fact biologically predisposed to trust in certain hinges.

I mentioned that this trust appears to be blind. Nothing seems to speak for its truth. This raises the epistemological question: Can fundamental epistemic trust nevertheless be epistemically justified? Is there an epistemically acceptable way to trust in cornerstone or rule propositions? Can it be epistemically good or right to do so?

I think it can. Under the auspices of Crispin Wright's 'Warrant for Nothing' (2004), philosophers have proposed different ways in which we may be entitled to accept our non-evidential beliefs. Entitlement is a form of warrant for which we do not need any epistemic achievement. I will critically examine the different ways that have been proposed to explain how we can be entitled to trust in a hinge proposition.

There are two broader strategies for generating entitlement: Consequentialist approaches and transcendental arguments. The consequentialist strategy argues that we are entitled to trust in hinges because it is the best means to acquire knowledge and true beliefs (Hazlett, 2006; Pedersen, 2020). The transcendental strategy is more ambitious, it argues that hinges are constitutively necessary for epistemic warrant, which entitles us to trust them (*cf.* Coliva, 2015).

In this chapter, I will examine how entitlement works. I first describe the *prima facie* epistemic status of hinges and trust in them. More specifically, I will argue they lack regular justification. I then propose a simple argument for why trust in hinges is nonetheless warranted, namely by entitlement. I also distinguish two kinds of criteria for something's being an entitlement: Positive conditions that are necessary for generating entitlement and negative conditions whose presence would defeat, taint, or reduce entitlement.

I focus on the positive conditions for entitlement. These describe the source of warrant. I introduce and critically discuss Wrightean entitlement. I reconstruct and explain the different arguments that Wright uses to introduce his account as well as his diverse arguments supporting entitlement because I do not want to presuppose knowledge of the debate. I then present a critical alternative internalist account of the normative state of hinges by Annalisa Coliva (2015). Finally, I introduce a different externalist account of entitlement developed by Tyler Burge (2003). The Burgean view, however, does not really fit with hinge epistemology, even though it can be considered as involving a notion of epistemic trust. Finally, I present my own view, *entitlement of cognitive activity*, which is a generalised transcendental account deriving from one of Wright's proposals.

## 2.2 The epistemic status of trust in hinges

I have described the doxastic role that trust in hinges plays. I will now examine its epistemic status. Can epistemic trust be justified somehow? As I have already argued in Chapter 1, any justification that I could get for a hinge would be circular. That is, any procedure or evidence that I could bring to bear on the justification of some rule or cornerstone would already rely on this rule or cornerstone as an implicit premise. Consequently, the hinge just established itself – which is a vicious circularity. This circularity undermines our hinges' ordinary justification and therefore hinges cannot be justified in the traditional sense.

As Lynch and Silva (2016) argue, even if you granted that circularity is *prima facie* unproblematic and a circular track record argument could generate justification for your hinge, the circularity would come back to bite us. On the one hand, circular track record arguments fail to serve as a reason and convince anyone who does not already share the hinge because they are circular (Lynch and Silva, 2016, p. 41), and why should an argument that cannot deliver convincing reasons be justificatory? On the other hand, circular justification fails given irresolvable deep disagreement. This is so because deep disagreement creates a tension between the fact that it is permissible to be conciliatory and change our position in its face, and the fact that deep disagreement gives us no epistemic reasons to change our position (Lynch and Silva, 2016, p. 47). I examine similar points in Chapter 3.

In what follows, I will show how this general argument works for different influential theories of justification. I show that hinges lack justification on five different accounts of epistemic justification. I begin with different foundationalist approaches: First, I examine evidential justification more closely. Then I consider internalist justification from a priori insight. Third, I examine externalist reliabilist justification. Fourth, as the last foundationalist straw, I consider dogmatism. Finally, I will consider whether hinges can be coherentistically justified.

### 2.2.1 Evidence

Hinge certainties are so fundamental that any other proposition that could be evidence for them already presupposes the truth of the hinge certainty. This is because hinges are singular propositions with far-reaching semantic and evidential consequences. They also demarcate the limits of our inquiry. We learn them not by insight or reflection but as brute facts. Some are innate, others are taken in as an aspect of how things are taken to be in our epistemic community. Sceptical arguments target the lack of evidential support for hinge propositions.

Not only is trust in hinges adopted without relying on any evidence, it also disregards contrary evidence. As I argued, we need to be at least as psychologically certain of hinges as any other beliefs for them to function as such. If you are certain that this here is an (external) hand, then you better also be certain that there are external objects.

- (C) A belief that P is a *certainty* if and only if it is held with such a degree of conviction that no other belief or evidence would change the degree of conviction with which it is held.

A belief that P is evidentially justified if and only if there is sufficient evidence available that speaks for P's truth. For illustrative purposes, we can



think of *sufficient* evidence as the amount of evidence required to make a true belief that P into knowledge that P under normal circumstances.

In the preceding chapter, I argued that cornerstone and rule certainties determine what counts as evidence for other propositions. Hence, any evidential support relation already presupposes the truth of some hinge. Therefore, any evidence for a cornerstone or rule proposition presupposes that it or some other cornerstone or rule proposition already is the case.<sup>1</sup> This ultimately produces a circular relation of evidential support between hinges. This is incompatible with justification because evidential justification is supposed to furnish *independent support* for a proposition. I do not think that agents are able to furnish an infinite regress of equally unwarranted hinges in support of the one in question. The set of hinges someone has available is arguably finite and thus if you tried to support hinges with each other, you would end up falling into circularity (*see also* Lynch and Silva, 2016).

As an illustrative example, consider THE BOOK containing the phrase ‘everything written in THE BOOK is true.’ *Prima facie* you might think that the fact that a source says that THE BOOK’s contents are true is evidence for the truth of THE BOOK’s contents. But given that this specific piece of evidence supports itself, that is, is circular, you already need to assume that THE BOOK speaks the truth in order for the evidential relation to kick in. This self-presupposing is incompatible with evidential justification.<sup>2</sup>

You might now argue that that’s all right but some other independent piece of evidence might support THE BOOK’s *bona fides*, maybe THE LITTLE BOOK. But you will hit the same difficulties with THE LITTLE BOOK. This means it is either self-supporting or it points back to some further piece of evidence and so on.<sup>3</sup> This is an example involving rule propositions.

To use a more realistic example, consider the cornerstone proposition ‘The world is older than me.’ If you want to find evidence for this claim, you will necessarily have to appeal to objects and testimony from people who are supposed to be older than you. For that to work, you have to presuppose that these objects and people are indeed older than you. Otherwise, the evidence would be merely apparent. This means any evidential support for your cornerstone already presupposes it and is not independent of it, hence, it is circular and does not generate justification.

Speaking in terms of hinges in general: You could support a hinge with some separate hinge, but this would be equally unsupported. Hence, you would either create an infinite regress of hinges successively justifying each other or fall into circularity by recurring to one of the previously justified hinges. Neither option is considered as yielding evidential justification because of the circularity or regress involved (Vogel, 2000), therefore hinge certainties cannot be justified by independent evidence.

### 2.2.2 *A priori*

The second kind of justification is internalist a priori justification. This is justification achieved independently of experience. There are different accounts of a priori justification. Some argue that it functions solely in virtue of grasping concepts as analytic a priori, for example (Ayer, 1952), others argue for some form of rational intuition, for instance (Bealer, 2000). There are also mixed forms. As you will see, this account somewhat straddles the other accounts because depending on your account of a priori warrant, it could be subsumed under evidential or a kind of reliabilist justification, which I treat in Section 2.2.3. However, as a priori insight historically speaking has played an important role as an account of justification for fundamental beliefs, I treat it separately here. These historical approaches tried to find *internalist* justification for our hinges: If our hinges are justified a priori, then we have epistemic access to their justification.

I will argue that hinge propositions cannot be justified a priori for the same reason as in the Section 2.2.1: The a priori insight would be circular, which undermines justification. Recall my preceding definitions of cornerstones and rule propositions:

- (CS) A proposition functions as a cornerstone proposition within a belief set if and only if the proposition is a necessary condition for a large set of other believed propositions. Cornerstones thereby determine the content of other propositions.
- (R) A proposition functions as a rule proposition within a belief set if and only if the proposition expresses a rule that is a necessary presupposition for rationally changing the belief set or rationally acting on the basis of the belief set. Rules determine how different believed propositions relate to each other as reasons.

One account of a priori insight comes from understanding concepts that yield analytic statements like ‘bachelors are unmarried eligible men’ or ‘bodies are extended.’ This conceptual insight, I would argue, relies on the cornerstones that determine the semantic relations between the concepts used in peripheral beliefs. There are central cornerstones for propositions about bachelors, namely some of the propositions entailed by all propositions about bachelors and all propositions about unmarried men. Thus, you might think: Great, I can infer the cornerstones from the more specific propositions about bachelors and support the cornerstones with that evidence. That is, we can interpret a priori conceptual insight as a form of evidential support.

This inference, however, already presupposes the cornerstone propositions which you inferred. As an example, take a set of specific bachelor beliefs you have: ‘John is a bachelor,’ ‘Mary is no bachelor,’ ‘Marc stopped

being a bachelor when he married,' etc. These bachelor propositions and your use of the term bachelor only have their specific content because you are already committed to a set of cornerstone propositions. You are already committed to the more central 'bachelors are unmarried eligible men' when you accept and form these specific bachelor beliefs. Therefore, the inference from the specific bachelor propositions to the more general cornerstone is circular. This circularity is incompatible with a priori justification of cornerstone certainties.<sup>4</sup>

This model of a priori justification could instead also be interpreted as an account of *self-evidential* justification. Namely, conceptual content delivers evidence for itself. Thus, conceptual insight delivers self-evidence (Burge, 1998). On this count, cornerstones are self-evident, given that they are the basis of analytic propositions.

However, this approach doesn't work for hinges either. There are two ways in which a proposition can be self-evident: The first is blatantly circular, every proposition entails itself, thus it is evidence for itself. That is not a live option for generating an a priori justification for hinges. The second model of self-evidence claims that understanding the hinge unlocks independent self-evidence for the hinge. This ideal may (or may not) work for mathematical and logical axioms, but it has an abysmal track record for claims about the world. Uncontroversially self-evident propositions like Descartes' 'I think therefore I am' are too weak to furnish proper hinges because there are not enough such Cartesian certainties to get a non-sceptical world view off the ground. Descartes (2013) already failed with respect to the mind-body problem and he had to resort to demonstrating God's existence to overcome scepticism about our cognitive faculties, circularly using those faculties.

The second account of a priori justification argues that it is the product of a capacity for a priori intuition. A priori intuition means that, by thinking about whether P is the case, you may 'feel'<sup>5</sup> that P is true, and this 'feeling' justifies the belief that P. By following an a priori intuition, you are relying on the rule presupposition that this 'feeling' does justify your beliefs as an implicit premise. This presages the hypothesis that hinges can be justified reliabilistically.

But why should you trust your intuition? Why should this 'feeling' be reliable, and how is it distinguished from similar but deceptive merely apparent intuitions? It is a rule proposition that you can trust an a priori intuition and the accompanying hinge as being reliable. Given its hinge nature, you cannot support this rule claim by evidence, but what other avenue do you have? You have to rely on the a priori insight itself to justify the rule that a priori insight is reliable. But this is methodologically circular, you rely on a rule to justify that very rule. This is incompatible with justification, which requires independent support. Thus, a priori intuition cannot deliver internalist, that is, accessible, justification for cornerstones and rules. Maybe it can deliver externalist justification for cornerstones?

### 2.2.3 Reliability

I now want to consider externalist<sup>6</sup> reliability in general as a third potential source of justification. A belief is reliabilistically justified if and only if it has been acquired by a reliable process or method. An example of this is the claim that I examined in the preceding section that a priori intuition could be a faculty that reliably delivers truths. If hinges are acquired through the reliable faculty of a priori insight, then it is reliabilistically justified. I will argue that beyond the issues considered earlier, the problem with this idea is that hinges are not really the kind of beliefs that can be reliably produced given their peculiar doxastic and epistemic position. The question is then what the scope of a reliable capacity is.<sup>7</sup>

A priori intuition is supposed to be a general-purpose faculty. It is supposed to be reliable not only for hinge propositions but also for many others, for instance, geometrical or logical propositions. Note that the aforementioned challenge that a priori justification is circular does not bother the reliabilist as long as the faculty is *de facto* reliable because it generates reliabilist justification in that case.

The trouble with a priori insight for hinges is that it does not seem to work. The rationalist project is considered as having failed. Fundamental propositions cannot be reliably discovered by a priori intuition as is evidenced by the fact that any and every fundamental proposition concerning the existence of god, for example, has already been defended on the basis of a priori intuition. If a priori intuition delivers such divergent results about hinges, clearly it is not a reliable faculty for hinges. On a side note, in statistics, ‘reliability’ just means that a process will produce similar results under similar circumstances; more recently, empirical research has shown that rational intuition in general is hardly stable and that it is susceptible to a wide range of extraneous, irrelevant factors (Rysiew, 2008). Therefore, it cannot be a reliable capacity.

May there be a less generalist reliable capacity delivering hinges for a certain domain? Different types of cognitive capacities appear to be reliable in different domains, for example, memory is reliable about the past and hearing is reliable about sounds in the organism’s immediate environment. Maybe domain-specific capacities can also reliably deliver true domain-specific cornerstones and rules?

I see a major obstacle for such a hypothesis lying in the fact that hinges are propositions so very different from the usual deliverances of such a capacity that there is no reason to think that it can deliver the former reliably. This holds especially given that domain-specific capacities usually depend on some kind of informational input, yet hinges are what is required to interpret information.<sup>8</sup> We do not remember or recall the hinge that memory is reliable and we do not hear that our hearing is accurate – memory and hearing are not rule-producing faculties.

Maybe there is a process that reliably delivers only hinges, thus justifying them? That is, perhaps we possess a hinge-producing faculty? I do not think we can think about the source of our hinges in the same way as we do about a reliable *process*. Regular reliable processes can in principle generate an infinite number of beliefs or at least a very large number. This is why it makes sense for them to be taken as reliable, they follow the law of big numbers. Hinges are different: An agent only needs a comparably small set of them, which is not subject to regular updating, and it seems to be odd to require that a faculty that produces such a small set of propositions be reliable. As a comparison, consider axioms in contrast to theorems: While it may make sense to postulate a faculty that is a reliable theorem prover, it seems strange to postulate a reliable axiom-finding faculty. The role that hinges and axioms play in a system of propositions makes them unamenable to reliability as a justifying property.<sup>9</sup>

The dedicated input analysers I examined in the previous chapter go together with an innate set of beliefs: Objects do not pass through each other, things that move autonomously are agents and so on (Carey, 2009, pp. 67, 157). Do we want to call the fact that organisms have these implicit innate beliefs a reliable faculty or process? I take core concepts to be orthogonal to reliability because they are not processes whose output is hinge beliefs; they are rather implicit hinge beliefs that are built into processes that produce other beliefs. They therefore cannot be justified by their reliability. Additionally, core concepts are not all veridical (Carey, 2009, p. 10) but are rather naïve beliefs. Hence, even if we overcame this orthogonality with respect to its *process* status, it is not clear that there is a *reliable* faculty there.

We do not get our cornerstone and rule certainties from some (reliable) process. Instead, we simply find ourselves with a set of certainties that are either built into our cognitive make-up or that come wholesale with our language and practices (Moyal-Sharrock, 2004). These two ways of hinge acquisition lack the form of processes that could be considered reliable.

#### 2.2.4 *Dogmatism*

We cannot get justification for our hinges off the ground through either evidence, a priori insight, or reliability. I will examine one last foundationalist option: Dogmatism. Jim Pryor (2000, p. 532), the godfather of dogmatism, argues that experience as of some proposition P gives you *immediate defeasible justification for P*. Note that this justification is not based on any evidence, reliability, or a priori insight. Instead, the brute fact that you experience P as being the case justifies the belief. The argument here would then be that you do not need to rely on any hinge because you are *immediately* justified.

Unsurprisingly, I do not think that our experience grants us any such immediate and basic justification. Namely, even this immediate justification from experience encodes a rule proposition that we can *prima facie* justifiably believe the deliverances of our experience as an implicit premise. We are committed to this rule, and if it were not somehow warranted, then we could not be warranted in the beliefs whose formation implicitly relied on the rule.<sup>10</sup> Note that we are unable to justify the rule and its cognates either through evidence, a priori, or reliabilistically.

I take Pryor to offer two considerations in support of his dogmatist foundationalism. The first is anti-scepticism: If we do not have foundational justification from experience, then we will fall victim to scepticism (Pryor, 2000, p. 519). Compare this to Van Cleve's (2003) argument for the goodness of epistemic circularity that I considered in the last chapter. In just the same way, I can argue that my proposal will give us a non-foundationalist way out, where we do not have to dogmatically thump the bible of our often unreliable experience. An anti-sceptic isn't forced to be a dogmatist.

Pryor's second argument for dogmatism is pointing out that it is *intuitive* that our experience justifies its corresponding beliefs (Pryor, 2000, p. 536). I do not think that this intuition gets his thesis off the ground. First, he would have to subsume his intuitions to his dogmatism – that is, argue that his intuition immediately and defeasibly justifies his dogmatism. But this is circular; we just used dogmatism to argue that dogmatism is true.

Second but relatedly, I do not think that people's intuitions about how experience supports our beliefs' epistemic status are as fine-grained as Pryor needs. This is because the intuition of immediacy Pryor appeals to is a fairly technical notion that aims to block a range of inferential, presuppositional, and evidential moves. Just as easily, I could elicit the intuition that we are committed to certain presuppositions about our perception when we form perceptual beliefs. This latter kind of intuition explains the mesmerising force of scepticism.

Finally, dogmatism needs to decide which experiences actually do justify immediately. Pryor does not think that all experiences furnish basic justification. Only propositions that are immediately presented in experience can be immediately justified (Pryor, 2000, pp. 538–539). This serves to exclude background presuppositions from infecting a belief's immediate justification. I would argue that in order for experience to even present propositions, background presuppositions, that is, rules and cornerstones, need to be operative – that is, for instance, what core cognition does.

According to Pryor, looking at a gas gauge indicating E gives, for example, no perceptually immediate justification for the proposition that the tank is empty (Pryor, 2000, p. 539). We can drive a wedge into this argumentative crack: What justification do I get from hearing an old recording? Should my seeing Sergey cry be immediate justification for my believing

that he is sad? What justification do I get from looking at Jupiter through a telescope? I do not think that there is a line to be drawn here. There are always presuppositions or implicit premises operative, even when I smell an orange that I am currently peeling. In all these cases, we can identify rules to which we are committed implicitly when forming our corresponding beliefs. Zoe Jenkin (2020) shows that even in seemingly simple cases of perceptual beliefs, core cognition and its associated hinges intervene to generate *mediate* justification instead of immediate justification.<sup>11</sup>

Note that I do not argue that we need to *infer* 'I have a P-experience, therefore P.' Rather, I argue that when you form a belief on the basis of your experience, you are committed to and guided by the rule that your experience is accurate, reliable, and occurs in normal circumstances as an implicit premise. This rule cannot then be justified by pointing to your normal, accurate, and reliable experience.

### 2.2.5 *Coherence*

There is one account of justification that I have not yet considered: Coherentist justification. According to this account, hinges might not be warranted by our entitlement to accept them, but by their *cohering* with all our other convictions. If coherence is an epistemic good that generates justification, then we do not need entitlement. Hinge certainties, by doxastically determining how other beliefs relate, will always cohere with these other beliefs. By always cohering with peripheral beliefs, our hinges will also automatically have a coherentist justification. That is, our hinges can never be incoherent with the rest of our belief set.

Having a coherent belief set apparently is an epistemic good, which becomes clear when we consider the criticisms to which incoherence is exposed. The interesting thing about coherentist justification is that circularity is a feature, not a bug in this view. The coherentist claims that justification *must* be circular because it arises from a belief's cohering with all the other beliefs. If we accept coherentism, we do not need entitlement because our hinges would already be justified by their coherence.

Given these considerations, one could consider epistemic entitlement as a sort of coherentism light. As will become clear, I do not take this to be the case; accounts of entitlement do not add any epistemic value to the supposedly already sufficient coherent justification.

I want to note that coherentism as an account of justification is about as contested as the idea of entitlement. At the same time, entitlement accounts are more economical, because they concern the warrant for a much more restricted class of beliefs than coherence is concerned with – namely hinges. Coherentist justification applies to each and every belief someone possesses.

Independently from its overriding epistemic entitlement, my problem with coherentist justification is that it has a strong relativist bent. The structure of coherentist justification is indifferent to how things are actually; a true coherent belief set is equally justified as a false one. Hence, the coherentist has to hope that her hinges are true, which is completely independent of how justified her account is. I will treat coherentism as a problem for entitlement more extensively in Chapter 3.

### 2.2.6 *The state of justification*

Is there then no justification or warrant to be had for instances of epistemic trust? Epistemic trust essentially aims at its own truth. It is not there to console us or to make life easier to bear, we possess it in order to come to epistemic grips with the world. That means if I trust in a false hinge, I would be better off if I stopped trusting in that hinge; and if I trust a true hinge, I am in a good position.

We can accordingly ask ourselves whether or when we are at least on the right track in trusting in the truth of a hinge. We cannot be epistemically justified in any traditional sense – so does that mean we always lack justification for our epistemic trust? That would be tragic: Given that all our other beliefs hinge on these propositions, they would themselves be unjustified.

Fortunately, this need not be the case. It has been argued that we may be *entitled* to trust in cornerstone certainties. This entitlement is a form of justification that we can have for propositions that we need to trust in, even in the absence of any traditional epistemic achievement.

## 2.3 Entitlement without cognitive achievement

I will begin with the abstract idea that entitlement is a kind of non-evidential epistemic warrant. That is, we may be warranted in accepting propositions without being justified by any evidence or any other cognitive achievement like reliability, a priori insight, or coherence.

This is a very general idea. I simply hypothesise that there is such an epistemic status that would give an appropriate warrant for fundamental epistemic trust. Up to now, I have neither established that there is indeed such a state nor have I given any specifics on how it would work.

The motivation for this general idea follows from the following simple consideration speaking in favour of the existence of entitlements:

### The Simple Argument

1. Everybody relies on hinge propositions in their epistemic practice (Hinge Epistemology).<sup>12</sup>



2. There are people who trust in hinges and who are also warranted in accepting and relying on some of their hinges (No Error Theory).
3. There is no epistemic justification to be had for hinges (No Justification).

*Therefore,*

4. There is a special kind of warrant to be had for hinges (Entitlement).

This argument sets some criteria for entitlement: From (Hinge Epistemology) and (No Error Theory), it follows that we may be entitled to some of the certainties that we actually hold.

In the background of (No Justification) lurks the assumption that regular justification is preferable to mere entitlement. This is plausible given that a state of entitlement has no non-circular support. This is different from regular states of justification where there is some independent state of affairs supporting the justified belief.<sup>13</sup> I will simply assume this.

THE SIMPLE ARGUMENT, however, is not incontestable. I take (Hinge Epistemology) to have been established in the preceding chapter and (No Justification) in the preceding section. But (No Error Theory) requires a defence.

One way of defending (No Error Theory) is anti-scepticism. Given that hinge propositions are the necessary presuppositions for all regular beliefs, unwarranted hinges threaten to undermine the justifications for regular beliefs. How can you be justified with respect to a regular belief if its necessary preconditions – the cornerstones and rules on which you relied for its justification – are unwarranted? Hence, the lack of warrant for our hinges implies that none of our beliefs are justified.

We can simply affirm that the sceptics are mistaken and that we are not in the deep epistemic predicament of lacking justification for all or most of our beliefs. By claiming (No Error Theory), I obviously beg the question against scepticism, because I affirm that some people have justified beliefs. But it only begs the question against the strongest form of scepticism, a global scepticism that denies justification for all epistemic states. (No Error Theory) leaves space for local scepticisms.

Note also the scope of denying (No Error Theory): Nobody is warranted in accepting any hinges at all. This means nobody is warranted in accepting, for example, the existence of an external world, that we have epistemic access to it, as well as all other hinges.

In any case, THE SIMPLE ARGUMENT is not there to defend entitlement theory as an anti-sceptical strategy. Rather, it appeals to non-sceptical philosophers by claiming they should consider entitlement theory as epistemologically important. Thus, (No Error Theory) is not that hard a sell to non-sceptics.

I think THE SIMPLE ARGUMENT has given us enough reason to explore how we could be entitled to trust in hinges. Before I introduce my own view, I will survey some of the positions on the market, notably the accounts defended by Wright (2004, 2014), Coliva (2015), and Burge (2003).

#### 2.4 Wrightean or internalist entitlement

I base my account of entitlement on Crispin Wright's (2004, 2014) account of epistemic entitlement, which is a response to the sceptical problem. Wright attempts to save internalist claims to justification from the sceptical worry. In this section, I lay out Wright's account of entitlement and compare it to competing views, while I will consider difficulties for the view in the next chapter.

He presents two kinds of sceptical arguments that can be traced back to René Descartes and David Hume, respectively. These arguments differ in their structure. One is based on considerations about underdetermination, and the other on considerations about closure.

Wright claims to have identified a gap in these sceptical arguments. Sceptical arguments undermine the claim that our cornerstone beliefs are justified. We may not have any (evidential) justification for our cornerstone beliefs, but there could nonetheless be some other kind of warrant that blocks the sceptical argument. He calls this entitlement.

Wright then introduces four possible ways in which we could gain entitlement to our cornerstone beliefs. These appeal to different kinds of considerations. Given Wright's great influence in the debate, I will first give a reconstruction of the argument I just sketched out. I will also explain how it has been received by other authors. This section of the chapter reconstructs and explicates Wright's different models of entitlement as well as the surrounding debate, if you are already familiar with the debate, you can also skip it.

There are two types of arguments supporting entitlement: Consequentialist and transcendental arguments. The consequentialist version of the argument focuses on the epistemic end state and argues that we are entitled because we maximise expected epistemic success as an outcome by trusting in hinges. This is an instrumental relation that moves from trust in hinges to epistemic success.

The transcendental version of entitlement, on the other hand, brackets the results and focuses on the epistemic activity itself. It argues that trusting in hinges is a constitutively necessary precondition for believing, inquiring, knowing, doubting, and so on. Without trust in hinges, we are epistemically and doxastically paralysed. Wittgenstein's argument from doubt (1969, §24) is an example of this strategy: The sceptic's doubts themselves already presuppose a commitment to hinges.

### 2.4.1 *Two sceptical arguments*

Wright's (1985, p. 431, 2004, pp. 168–169) Cartesian argument uses *sceptical scenarios* to undermine our ordinary beliefs. Basically, a belief that P is only justified if my evidence excludes all alternatives compatible with the current evidence. My belief that I just saw a badger running past me in the dark is only justified if my evidence also excludes the alternative that a cat or a dog ran past me.

The sceptic raises alternative possibilities that we cannot exclude through an appeal to evidence. The simplest is the scenario that I am currently dreaming (Descartes, 2013). Any procedure I undertake to gain evidence for my not dreaming could be dreamt. Given its structure, the sceptical dream scenario prevents us from ruling this possibility out by an appeal to evidence. No evidence is capable of establishing that I am not currently dreaming because, if I were dreaming, I would also be dreaming this evidence. We therefore lack justification for the belief that we are not dreaming as well as for our other beliefs.

As you may have realised, the proposition that you are currently not dreaming is a hinge. It is a more specific instance of the rule that our sense experience gives us information about how things are. There are many alternative sceptical scenarios, from deceiving demons to evil scientists who have trapped your brain in a vat.<sup>14</sup>

Wright's (1985, p. 434, 2004, pp. 169–172) Humean sceptical argument attacks Moorean reasoning (Moore, 1939), which latter has a three-step structure. This sceptical argument first presents an instance of Moorean reasoning and then points out where it fails. Finally, it explains why the failure can be generalised to cover all beliefs in a given domain.

The Moorean Argument aims to establish that there is indeed an external world. It goes as follows. Given some proposition P about the external world, for instance, 'The cat is sleeping on the mat':

- I I have evidence to the effect that P (I have an experience of the cat sleeping on the mat.).
- II Therefore P (The cat is sleeping on the mat.).
- III Therefore, an external world exists.

The inference from II to III may seem strange, but it is a simple entailment. III is a necessary condition for II, given that P is about an external world. If there are cats and mats, then there is an external world. Therefore, if P is the case, then an external world exists.

The problem in these arguments, however, is rather the inference from I to II. As mentioned, III is a necessary condition for II. If I want to be justified in inferring II from I, then I need to already assume III. This means that for the argument succeed, its conclusion must be foregone.

III is a hinge proposition and, as I have argued earlier, all possible arguments in favour of III are of this I-II-III structure. Given this circularity, it is not possible to gain independent justification for cornerstone propositions like III. We can apply this reasoning to any cornerstone proposition. There will be no evidential warrant to be gained for it, because in inferring an evidential justification for a cornerstone, we already presuppose the cornerstone. Hence, this argument can also be generalised to cover other claims, for example,

- I Karola grimaces when something touches her elbow and she avoids moving this joint.
- II Therefore, Karola is in pain.
- III Therefore, there are other minds.

We have to have already assumed that there are other minds to infer that II 'Karola is in pain' from her behaviour I.

These are the two sceptical problems that Wright raises.<sup>15</sup> Each functions by showing that we lack internalist warrant for our cornerstone beliefs. If our cornerstones are not justified, then this lack also includes other beliefs for which these cornerstones are the necessary conditions.

The relationship between these two arguments has seen some debate, especially by Pritchard (2016, pp. 54–55), who argues that these two arguments are so different that they require separate treatment. I will stick with Wright's (1985, p. 438, 2004, 2014) and Coliva's (2015, 2021) one-size-fits-all approach and not go into further detail here given that the correct analysis of scepticism is secondary for this project.

#### 2.4.2 *The gap*

Wright (2004, pp. 174–175) observes that there is a gap in both these arguments. What the two sceptical arguments show is that there is no evidential justification to be had for these cornerstones as any evidential justification would rely on previously assuming justification for the cornerstone beliefs at issue.

However, for the sceptical argument to hold, the sceptic needs an auxiliary assumption: Evidential justification is the only kind of justification we can have for cornerstones. Wright denies this assumption.

Wright (2004, p. 175) suggests that we may be *entitled* to claim these cornerstones and rules are true. Entitlement here means a sort of default warrant which does not need to be achieved cognitively. When you are entitled to trust your hinges, you need not worry about scepticism anymore and can happily achieve justification for your ordinary beliefs. This is the same gap which THE SIMPLE ARGUMENT used to make its point.

This proposal for entitlement raises the spectre of circularity again: Once my entitlement enables me to get a justified belief that there is a hand, I can competently and in Moorean fashion infer that there is an external world. I can then justifiably believe the hinge proposition that there is an external world.

This problem has been aptly named *epistemic alchemy* (Davies, 2004, p. 222; Wright, 2014, pp. 229–230) because the lead of entitlement is alchemically transformed into the gold of justification. Wright (2014, pp. 233–234) permits this kind of circularity, but only conditionally on your first having gained entitlement. Alchemy then does not improve your epistemic situation (McGlynn, 2014).

I am also willing to grant that justification transmits to your hinges; but it does nothing with your *doxastic* status. Recall, that you are already certain in your hinge trust; the evidential justification that you get from your experience *does not improve your situation*. It just adds a justificatory coat of gold to your hinge trust rather than transforming it into an evidentially justified conditional certainty of massive gold.

Usually, justification is bound up with some cognitive achievement: Successful acquisition and evaluation of the evidence, a successfully executed process or method, or an a priori insight (*cf.* Sosa, 1991). Entitlement is suggested to be a warrant we can possess without any of those. We just have it when we accept a proposition as true. While being entitled does not require any cognitive achievement, it is not unconditional (Wright, 2004, p. 175). The cognitive achievement generates the normatively good state of justification. So, what *generates* the normatively good state of epistemic entitlement?

In the broader literature, there are two kinds of conditions that determine whether we are entitled in accepting a proposition: Negative conditions and positive ones. The negative conditions defeat the epistemic good of entitlement, thus their absence is a necessary condition for your being entitled. Contradictory hinges, for example, arguably undermine your entitlement to accept either of the two hinges. Hence, negative conditions are defeaters for entitlement.

Positive conditions for entitlement do the epistemological heavy lifting (Wright, 2014, p. 221). They give the formal conditions for those occasions on which we have an entitlement, barring the emergence of some negative condition. The positive conditions for entitlement need to explain what the source of epistemic entitlement is and why entitlement is a form of epistemic warrant.

This means that in order for there to be epistemic entitlement, it is necessary that at least one positive condition for entitlement holds and that there is no negative condition. My own position, that I lay out in Section 2.6, is that hinges are necessary preconditions for cognitive activity aimed at truth, and this is the positive condition for entitlement. I will focus more on negative conditions from Chapter 3 onwards.

### 2.4.3 Entitlements

I will bracket the negative conditions on entitlement for now. Suffice to say that they usually serve to guarantee theoretical and practical consistency, and rationality, and to avoid a diverse set of difficulties. I will instead focus on the different arguments that epistemologists have available for our being warranted without cognitive achievement according to Wright. That is, I present the different positive conditions for entitlement that Wright proposes, but I also point out where they fall short and where they need further clarification.

#### *Reichenbachian entitlement*

The first approach, which Wright (2014) defends more extensively, is so-called *strategic entitlement*. According to Wright (2004, p. 179, 2014, p. 224), trusting cornerstones is an epistemically dominant strategy. Therefore, we are entitled to trust in them. This strategy is derived from Reichenbach's (1938) solution to the problem of induction, which in its turn is inspired by Pascal's Wager. As you will quickly recognise, this is clearly a consequentialist form of entitlement.

Strategic entitlement is gained by considering whether we should trust some particular cornerstone C or not. The cornerstone may be either true or false, which leads to four possible outcomes.

Here, trusting in C is the dominant strategy. Only if we trust in C do we have the chance of gaining any true beliefs. All the other cases are equally bad; thus, the most rational strategy is to at least attempt to gain many true beliefs by trusting in C. It would be unreasonable and unduly pessimistic to remain a sceptic about C, because you should at least take a shot at getting things right. Therefore, you are entitled to trust in C. This is an instance of consequentialist reasoning.

Quite a few objections have been made to this strategy, usually challenging consequentialism's epistemological pragmatism under which you would trust your hinges for the wrong reasons; for example, if a fairy offered you lots of knowledge if you believed a particular proposition without any evidence (Pritchard, 2014a; Elstein and Jenkins, 2020). Also, the

Table 2.1 Decision table for trusting a cornerstone (*cf.* Wright, 2014, p. 225)

	<i>C is true</i>	<i>C is false</i>
<b>Trust C</b>	MANY TRUE AND USEFUL BELIEFS	Few true and useful beliefs
<b>Do not trust C</b>	Few true and useful beliefs	Few true and useful beliefs

usual objections that can be made against Pascal's Wager (James, 2014) – the frivolity of an epistemic gamble<sup>16</sup> and its being swamped by additional bad alternatives could be raised with some appropriate modifications. I will not review all of them here. I will instead focus on a single criticism that I take to be fatal to strategic entitlement because it engages with the epistemic consequentialist on their own terms.

Wright's argument relies on the assumption that trusting in a true cornerstone is good while all other possible outcomes are equally bad. Pedersen (2009, 2020) criticises this assumption. Clearly, if we do not trust in C, the outcome is the same independently of whether C is the case or not. The key question is what trusting in a false C implies and how good or bad it is as an epistemic outcome.

The problem is that if you trust in a false cornerstone C, you not only have few true beliefs but most likely also have many false beliefs. By trusting in C, you also accept many beliefs. If only a few of those are true (as Wright accepts), then the rest must be false, assuming bivalence. In general, having false cornerstones and rules would arguably lead to a proliferation of false beliefs. You would tend to rely on unreliable methods and mis-categorise objects on the basis of your rules and cornerstones.<sup>17</sup> Avoiding error is clearly an epistemic good, a position that has a venerable tradition behind it (Clifford, 1999; Descartes, 2013).

This muddies the waters for Wright's argument: The Reichenbachian now needs to additionally argue that the potential prize of getting things right outweighs the potential cost of being wrong. Arguably, this cannot be established (Pedersen, 2020).

The easiest way for Wright to get his argument to work would be to claim that truth has more value than falsehood has disvalue. Assume for argument's sake that the value of getting things wrong is neutral, as in Wright's table. Then the prize of getting things right would simply outweigh the risks, which would be nonexistent. If this were the case, then it would be reasonable to always simply believe both a proposition and its negation. It would at least be strategically reasonable. The value of getting it right would (massively) outweigh the cost of getting it wrong. This seems false. For that structural reason, I think that a true belief's false negation must weigh up the true belief's epistemic value in epistemic disvalue.

*Table 2.2* Decision table for trusting a cornerstone (*cf.* Pedersen, 2009, p. 450)

	<i>C is true</i>	<i>C is false</i>
<b>Trust C</b>	MANY TRUE, FEW FALSE BELIEFS	<i>Few true, many false beliefs</i>
<b>Do not trust C</b>	Few true, few false beliefs	Few true, few false beliefs

You might object that believing all propositions would be irrational on other counts, namely because you would believe many contradictions. This point is mooted by the fact that belief in contradictions is usually prohibited because it leads to explosion (*ex falso quodlibet*). In a belief set containing every proposition, as just described, the explosion from inconsistency does not make you epistemically worse off. What remains are the supposedly better true beliefs that have surplus value with respect to their false counterparts. Something has gone wrong if believing everything becomes the dominant strategy on this account. Believing everything is in many ways similar to believing nothing, because neither approach distinguishes between alternatives.

If true beliefs were more valuable than their false counterparts, an alternative dominant strategy would be to arbitrarily believe any proposition that occurs to you. The potential gains of getting it right would outweigh the threat of getting it wrong. Therefore, false beliefs, including trust in false cornerstones, make you as worse off as getting things right would make you better off (Pedersen, 2009, 2020).

Given these considerable difficulties, I will disregard strategic entitlement. I do not think that epistemic consequentialist accounts of entitlement can get off the ground.<sup>18</sup> My account is based on the model I present in the following section.

#### *Wittgensteinian entitlement*

*Entitlement of cognitive project* derives from the Wittgensteinian project. It is inspired by two passages in *On Certainty* (Wittgenstein, 1969, §§163, 337). Given its proximity to my own project, I will focus on this approach. It is also the best worked-out framework on the market. As will become clear in the course of this section, entitlement of cognitive project can be read either as a transcendental or as a consequentialist account.

It is important to note that Wright (2004, p. 190) does not argue for entitlement simply by virtue of the status of hinges. Entitlement of cognitive project is not focused on hinges, rather it examines the structure of investigation within a particular context.

Referring to Wittgenstein, Wright (2004, p. 189) notes that any investigation relies on a set of presuppositions that are necessary for it to follow through successfully. Wright (2014, p. 215) also calls them ‘authenticity conditions’ for an investigation. These presuppositions need not be as fundamental as hinges. One can verify them. For example, if you investigate how full your tank still is, you presuppose that your fuel gauge is functioning. You can verify that independently by having the gauge checked. This verification, however, will also come with a set of presuppositions which you need to verify. In a sense, having verified your first presupposition does not put you in a better epistemic



position concerning the presuppositions for your cognitive project (Wright, 2004, pp. 190–191). If you doubted the presuppositions, then the cognitive project would be meaningless (Wright, 2014, p. 215).

Jochen Briesen (2012) has developed a detailed account on how presuppositions work on this approach by introducing philosophy of language. He points out that cognitive projects essentially involve the answering of a question. The semantics of a question not only involve the explicit question itself; the question is asked with a set of implicit presuppositions.

For example, a presupposition of every question is that the entities or properties asked about do exist, except if you ask whether they exist. If the presupposition is false, the question is bound to be rejected as nonsensical or as a trick question. Take, for example, the question: ‘Are lerts pink?’ The question is invalid because there is no such thing as a lert and the statement ‘I am a lert’ is meaningless. However, if you investigate whether lerts are pink, then you simply presuppose their existence – as we can see from examining the corresponding question.

If you use the requirement that the presuppositions of every cognitive project must be verified by some independent project as an epistemic standard, then this generates an infinite regress of justificatory projects and questions that need to be answered. Doubting every presupposition makes answering any question and completing any cognitive project impossible (Wright, 2014, p. 216).

The source of entitlement of cognitive project is that presuppositions are a necessary precondition for investigation. Here, the distinction between consequentialist and transcendental entitlement becomes apparent: Are we entitled because of the potential results of investigation or because it enables investigation *per se*?

This distinction tracks a debate about presuppositions in the philosophy of language: Sentences like ‘the current king of France is bald’ or ‘lerts like cookies’ are either taken to be false or meaningless (Beaver, Geurts and Denlinger, 2021). If they are taken to be false, this follows the consequentialist reading: A false presupposition simply means the investigation fails. But if the sentences are taken to be meaningless due to false presuppositions, then cognitive projects with false presuppositions will also be meaningless and not real investigations. On this count, the presuppositions constitute the investigation, which is a transcendental argument.

On the consequentialist reading, investigation is valuable because we cannot acquire knowledge without active inquiry. This means presuppositions are a means to the valuable end of knowledge, which would otherwise remain inaccessible.<sup>19</sup> This means–ends relation is the consequentialist source of entitlement of cognitive project and it explains why we are warranted in trusting in presuppositions or cornerstones. We would be unable to gain any knowledge without them.

On the transcendental reading, the investigation itself is taken to be the source of value. Because the pursuit of cognitive projects is a good and because we are committed to this good, we are entitled to trust in the presuppositions of our investigations. The presuppositions are what render cognitive projects possible. Not trusting them would end in ‘cognitive paralysis’ (Wright, 2004, p. 191).

This transcendental account of entitlement also gives us a response to the challenge of the truth fairy (Wright, 2014, pp. 241–242; Elstein and Jenkins, 2020) that I hinted at in the subsection on Reichenbachian entitlement: Imagine a fairy offered you a lot of knowledge if you accepted Goldbach’s conjecture without any evidence – this would give you consequentialist entitlement but accepting the conjecture would compromise your epistemic values. However, accepting a hinge to enable you to even be epistemically active is no such compromise – there would be no epistemic values to pursue without trust in hinges. Coliva (2020, p. 340) thinks that Wright does not consider entitlement to be transcendental, I believe Wright is uncommitted.

I do not think that this entitlement must remain limited to the rule presuppositions of particular cognitive projects. The argument can be harnessed for a more global kind of entitlement. I will argue in Section 2.6 that hinge certainties in general function like presuppositions of cognitive project: They are the necessary conditions for our cognitive activity in general. If you wanted to verify a hinge, the only way to do so would be by proffering a further hinge of equally uncertain standing (Wright, 2004, p. 192). Hence hinges are the most general presuppositions. This is the exactly same justificational structure as the presuppositions for cognitive projects and thereby for the entitlement of cognitive project. I shall argue for the entitlement of cognitive activity as a generalised version of entitlement of cognitive project. Wright (2014, p. 216) also works towards such a generalisation but his approach is to argue that many particular cognitive projects share their presuppositions.

### *Strawsonian entitlement*

This account of a source of cognitive entitlement is broadly Kantian. Wright (2004) only sketches out this transcendental argument for why we are entitled to accept a mind-independent reality. The argument was inspired by considerations drawn from Strawson (1959) and Evans (1985), which examine what sort of experience could warrant the notion of a mind-independent world.

Assuming that conceiving of experience as objective is somehow independently mandated . . . a somewhat minimal notion of entitlement of substance might then emerge: since some conception of one’s cognitive

locality and of the substance of states of affairs that are elsewhere is essential to any objective conception of experience – and since (suppose) so conceiving of experience is independently warranted or unavoidable – a thinker is entitled to the basic ontology involved in an otherwise coherent conception of what kind of thing might obtain at other localities.

(Wright, 2004, p. 203)

As far as I understand this line of reasoning, Wright posits that

- A Essentially, experience presents itself as about a mind-independent reality.<sup>20</sup>
- B This is not coherently possible without our implicitly accepting the cornerstone that there is a mind-independent reality or – to speak in Kantian terms – that there are things in themselves.
- C We are therefore entitled to accept that there is a mind-independent reality.

If this reconstruction of Wright's suggestion is correct, then we are entitled by a *conceptual schema* (Evans, 1985) to accept the existence of a mind-independent reality. This entitlement is reached through a classic transcendental deduction and comes with the difficulties that accompany such reasoning.

This approach does fit well with my own account of entitlement. Notably, it accords with the notion of core cognition; that is, the innate conceptual framework we are equipped with. This account of entitlement explains entitlement of fundamental cognitive activity, that is core cognition.

However, it has some difficulties as an anti-sceptical strategy. Most notably, the sceptic could simply reject A and argue that experience presenting itself as representing a mind-independent reality is a contingency given that it is *possible* that there is a Berkeleyan being who experiences things as a mere sequence of impressions and nothing else.

For the transcendental argument to work against the sceptic, we would need to show that the notion of experience becomes incoherent without it. I do not think that there has been a truly successful argument against Berkeleyan idealism, that is, a conception of experience being all there is.

Putnam's *brain in a vat* arguments (Putnam, 1981) come the closest to offering a rebuttal to such an idealism. But what they show is that *our* experience as it actually is could not be had in a mere dream of a brain that has been envatted for all of its life. This point thereby strengthens point B in our argument. However, even Putnam's argument has its limitations: If we knew of people who just dreamt their entire life (e.g., patients comatose from birth), then we would have to grapple with the possibility that we may be among them, thereby undermining the argument.

As was mentioned earlier, for the purposes of my project, I am happy enough with this state of affairs. Wright himself openly admits that this line of argument has many difficulties and that it is not yet sufficiently developed to stand on its own (Wright, 2004, p. 202). But while it may not convert the convinced sceptic, it shows her a way out: It tells the sceptic and everybody else that they are entitled to trust in their conceptual schema because they happen to have this conceptual schema and other schemas are hard to even imagine.

As a last note: On its own, this kind of entitlement risks implying that we win the battle but lose the war against the sceptic. I hinted at the argument's Kantian roots, namely, it is an argument to claim entitlement for accepting the existence of things in themselves.

The problem here is that entitlement to accept a mind-independent reality does not yet mean that we are entitled to claim epistemic access to this mind-independent reality. Kant (1998) himself famously thought that this is not possible. Hence, this entitlement from conceptual schema cannot stand on its own as an anti-sceptical strategy. Even though we might be able to argue that our experience presents itself as granting access to external reality and that we are unable to conceive of it otherwise; the problem of perception, that is, the fact that we could dream and hallucinate in the very same way becomes even more acute here.

#### *Entitlement of deliberation*

The last argument uses the notion of *rational activity* to generate entitlement. Namely, if you want to act rationally, you need to deliberate; which comes with a whole set of presuppositions. We cannot establish presuppositions, therefore we may as well trust them to hold, thence the entitlement.

The generic thought is that since rational agency is nothing we can opt out of, we are entitled to place trust in whatever (we have no evidence against and which) needs to be true if rational decision-making is to be feasible and effective.

(Wright, 2004, p. 198)

This is, in a way, the practical counterpart to entitlement of cognitive project. It does sound like a pragmatist account of entitlement and hence of epistemic warrant. But this characterisation is not quite correct and I do not think that Wright cares about action *per se*. Rather, he is interested in our rational agency in general, independently of specific action. The semblance of pragmatism results from the fact that rational agency principally manifests itself in practical activity. Bratman (1992) argues that 'acceptance' forms the cognitive background to our practical agency, thus it is plausible that Wright found some (unrecognised) inspiration within his approach.

If this non-pragmatist interpretation is correct, then Wright must consider rational agency as being a source of epistemic value. You could subsume it under entitlement of cognitive project. The relevant cognitive project is finding out what you should do.

Wright would probably not want to collapse rational deliberation into cognitive projects. In that case, there seem to be two options: go down a pragmatist path, which I take to be anathema to epistemic warrant or argue that rational deliberation is an epistemic good *sui generis*. The latter idea could be subsumed under the account of entitlement from cognitive activity that I will propose. Wright clearly demarcates entitlement of rational deliberation from consequentialist strategic entitlement (Wright, 2004, p. 200), thus I take his argument to be transcendental, advocating hinges as constitutive of the good of practical agency.

James Foley (1991) proposes a similar view, in which we reasonably accept certain propositions without evidence. Annalisa Coliva also argues that we need our hinges because they constitute rationality. However, she extends rationality beyond practical considerations. She thus argues that we are transcendently required to trust in hinges in order to be rational. However, she does not think that this gives us epistemic entitlement.

#### 2.4.4 *Extended rationality*

I distinguish entitlement from the question of whether trusting in hinges is rational. The rationality of an epistemic state is a many-headed beast. There may be a sense of ‘rational’ which coincides with the epistemic justification I am asking for, but at the same time, there are many other senses of ‘rational’ which have little to do with this. For example, there is a minimal internalist sense of ‘rational’ that just requires that your belief set be consistent. Another sense concerns whether or not you follow your reasons. What these reasons are is left completely indeterminate.

Annalisa Coliva has developed a proposal that is structurally similar to Wrightean entitlement but is an independent position. The similarity is that it is also committed to a form of epistemological internalism. Additionally, the arguments she uses have a similar structure to Wright’s. Meanwhile, she does not think that entitlement as a form of epistemic warrant is possible.

Coliva argues that epistemic warrant must essentially be epistemically good. Otherwise, it could not be *epistemic* warrant. Further, she argues that the only way a state can be epistemically good is if it bears some indicative connection to how things are. A doxastic state is epistemically warranted if and only if there is some available indication that it is true.<sup>21</sup> This is what the ‘epistemic’ in epistemic warrant is supposed to mean: Epistemic warrant is evidential warrant (Coliva, 2015, p. 64 ff., 2022, p. 175). I challenge this evidentialist assumption at the end of this section.

This means that epistemic entitlement cannot be a form of epistemic warrant as Coliva understands it. The entitlement theorist concedes as much to the sceptic: We cannot find any good evidence for our hinges. We have no non-circular indication of their truth. If a direct indication of truth is required for warrant, then entitlement cannot get off the ground.

According to Coliva, we therefore cannot be warranted or entitled to accept cornerstones. This, however, leads to a different problem: Are we making a mistake in accepting our hinges? At the least, trusting our hinges is unwarranted. Some would argue that, because of the lack of epistemic warrant, it is also epistemically *irrational* to trust in hinges as, the argument goes, epistemic rationality simply means sensitivity to epistemic reasons.

Coliva resists this move. She argues that we are mistaken if we suppose that rationality simply coincides with epistemic warrant. That is, she does not limit rationality to a mere sensitivity to epistemic reasons – evidence. Instead, she advocates a notion of *extended rationality* (Coliva, 2015, p. 119).

Rationality is not merely conditional on a belief system. Coliva (2015, p. 146) uses the metaphors of the rules of a game: One way to be rational is by following its rules. That is the conditional sense; you are rational if you follow the reasons you have.

The thesis of extended rationality is that there is more to rationality than following rules: *Accepting them* is part of rationality. Namely, there are some propositions that constitute the game of rationality. Their acceptance is *extendedly rational*. Some propositions are necessary to gain epistemic (or perceptual) warrant, and we cannot get any warrant for these (Coliva, 2015, p. 129) because they are the preconditions for epistemic practices. I call these propositions hinges. Accordingly, on the extended rationality view, it is (extendedly) rational to trust in hinges, but you are not epistemically warranted to do so. This is a classic transcendental argument: Trust in hinges is constitutive of rationality, therefore it is rational to trust in hinges.

This may sound disappointing as an anti-sceptical strategy. It sounds nice to know that we are not irrational in trusting hinges, yet we are not epistemically warranted in doing so either. Is that good enough?

It is, insofar as it is highly efficacious as an anti-sceptical argument. If the extended rationality thesis is correct, then the sceptic is irrational in rejecting hinges. She does not accept all the propositions she would be rationally required to accept (Coliva, 2015, p. 129). If you do not accept these hinges constitutive of rationality, you cannot be rational. In a way, by rejecting her hinge beliefs, the sceptic opts out of the game of rationality without realising it.

As mentioned, I consider this to be an internalist view, given that it does not appeal to any external factors. I therefore group it together with Wrightean entitlement views, even though it is distinct from epistemic entitlement.

According to Coliva, entitlement views ask for too much: Epistemic warrant requires a truth connection that is unavailable for cornerstones.

Coliva's metaphor of the rules of a game makes two issues salient. First, which game should we play? Second, what if I do not want to play anymore? I will confront the first issue, namely that there may be different rules constitutive of different rationalities, in the next chapter with respect to the threat of relativism.

Coliva calls the second issue the 'Oblomovian challenge' (Wright, 2014, p. 244; Coliva, 2015, p. 145). It can be summed up in the question 'Why be rational?' (Kolodny, 2005). Coliva argues that hinges are constitutive of *epistemic rationality* and even asking for reasons why you should be rational is engaging in the exercise of rationality, and this requires hinges. The extended rationality view creates its own reasons for itself. Thus, Coliva follows Wittgenstein's (1969, §24) manoeuvre concerning doubts.

The entitlement theorist, however, will argue that indirect possible access to truth, rather than direct indication of truth, is sufficient for entitlement. That is, entitlement theorists, *pace* Coliva, argue for an *extended warrant* thesis. They point out that the project of coherent epistemic warrant cannot get off the ground without entitlement (Wright, 2014, p. 243). They would say that Coliva falls into the same mistake as the sceptic, namely, believing that epistemic warrant can only be evidential warrant.

Can non-evidential entitlement then warrant your trust in your hinges, even though it does not evidentially support them? Like Wright, I am inclined to think that yes. Although my hinges are not evidentially justified, there is this broader epistemic warrant called entitlement. Note that while entitlement is not warranted by evidence, it is nevertheless supported by the epistemic reasons – rehearsed in this chapter – that speak in favour of accepting a hinge. These reasons give us epistemic entitlement but not evidential justification. I do not see why these considerations about our ability to gain knowledge, understand the world, be cognitively active, etc. should not be epistemic reasons. I will further expand on the epistemic normativity underlying entitlement in Section 2.6 of this chapter.

## 2.5 Burgean or externalist entitlement

*Prima facie*, Tyler Burge's (2003) externalist account of epistemic entitlement has some similarities to Wright's account. Notably, Burge also distinguishes between two kinds of warrant: Evidential justification and reason-independent entitlement. Both accounts of entitlement aim to grant warrant without epistemic achievement. Is there a way to bring the two together?

Hardly: While, for Wright, entitlement is something of a last epistemic resort, for Burge, entitlement is the default. According to Burge, we have lots of beliefs for which we do not cite reasons but which are nonetheless

warranted. They are entitlements. Notably, perceptual beliefs are entitlements in normal environments because there is an evolutionarily generated process, perception, that is reliable in the environments that produced it. But there are no reasons that we can cite in favour of them (Wright, 2014, p. 223).

Reasons are accepted propositions that support a belief that, according to Burge, you need to believe if they are to justify another belief. A justified belief is a belief supported by another believed proposition. In Wright's book, a belief is justified by epistemic achievements, while entitlement is supported by non-evidential reasons – namely transcendental or consequentialist considerations. Hence, Burge and Wright disagree about what extensionally counts as justification and entitlement (Graham et al., 2020, p. 7).

The problem with justification, according to Burge, is that it is inaccessible to children and animals who are unable to hold propositional beliefs or to relate them. This would prevent them from ever being justified. On this account, they would never be OK, epistemically speaking. Such a hyper-intellectualisation is absurd – and we therefore need a further kind of non-propositional warrant called entitlements (Burge, 2003, p. 505).

Essentially, Burgean entitlements are the warrants we possess for beliefs that have been reliably produced by our perceptual or cognitive faculties. Hence, entitlement is an externalist reason-independent warrant. Given that both children and animals possess reliable faculties, this is the kind of warrant that they can gain. If not, epistemic warrant would risk being hyper-intellectualised (Burge, 2003, 2020).

Thus, for Burge, every perceptual belief is an entitlement. We do not need any reasons to be warranted in accepting them.<sup>22</sup> However, entitlement and justification are not mutually exclusive. You may find also reasons for something to which you are entitled. It simply reinforces the warrant.<sup>23</sup>

This is markedly different from the Wrightean notion of entitlement. According to Wright, we are only entitled to propositions for which it is not possible to obtain regular non-circular evidential justification (Graham et al., 2020, pp. 7–8).

Can the two views nevertheless be brought together? Strictly speaking, no, given that the underlying motivations are so divergent – one is about perceptual faculties, and the other is about the structure of warrant. We can nevertheless examine what rules and cornerstones do in the Burgean framework.

Burge (2020, p. 90) thinks that his perceptual entitlement is all we need. Once we have perceptual entitlement, we can then support our cornerstones with reasons. For example, our perceptual entitlements entail that there is an external world.

In a way, the rules and cornerstones are embedded in Burge's perceptual entitlements. They are built in by externalism. We are entitled, because there is an external world and because of the reliable processes that have been selected for through evolution.



So while the two accounts cannot be made to work together, they have some interesting parallels. If you described an epistemic subject's internal perspective on Burge's account, it would be strikingly similar to a subject's point of view on a Wrightean account. Both the Burgean and the Wrightean subjects rely on their senses' reliability. Recall that, on my view, there does not need to be an explicitly held hinge certainty for there to be an entitlement. Rather the hinge is implicit in the agent's epistemic behaviour and the agent is entitled to behave in this way. This is especially relevant for my account, which argues that *core cognition* consists of hinge certainties.

## 2.6 Entitlement of cognitive activity

I developed THE SIMPLE ARGUMENT to argue that we should try to find an account of how we can gain entitlement for trusting in hinges. Note that here I will only provide positive conditions for entitlement. That is, these conditions explain why we are entitled to trust in such propositions – what the source of entitlement is – but they are merely necessary conditions.

There are also negative conditions for entitlement whose absence guarantees that the entitlement is not defeated. I will treat those in Chapter 3. But it can already be noted that I do not think that one of the defeating negative conditions is that there be no counterevidence, as Wright (2004, p. 191) thinks. Given that hinge certainties determine semantic and epistemic relations, there are no defeating evidential reasons against a hinge to be had on my view.

The source of warrant with cognitive entitlement, I want to claim, is cognitive activity. Not just any cognitive activity, but cognitive activity aimed at finding truth. By this, I mean the search for truth in the broadest sense. I argue for this via a transcendental route: We are entitled to trust in hinges because they are the constitutive conditions for searching for the truth. Nikolaj Pedersen (2009) sketches out a similar proposal: The source of entitlement is grounded in its teleological value, which can be realised independently of its ultimate success.

You might wonder whether any cognitive activity whatsoever is a sufficient good to allow the obtaining of an entitlement. The reason that this is not so is simply that not all cognitive activity is inherently truth-directed. Dreaming is a (mostly involuntary) cognitive activity which apparently does not aim at truth in any conceivable sense for us. Spinning some yarn and making up a story is also a cognitive activity, but is again one that does not really aim at truth. As a last example, consider Escher drawings: They arose from cognitive activity but explicitly try to represent the impossible. Given that such cognitive activity is detached from how things are, including in what it aims at, it does not generate epistemic warrant.

My thesis is then that truth-directed cognitive activity is an epistemic good. I shall call this the *search for truth* or *cognitive activity* for ease of expression. This is in a way a more generalised formulation for the account of entitlement of cognitive project. But rather than tying entitlement to single investigations, I link it to investigating in general.

I am inclined to think that the search for truth is an epistemic good. Thus, while actually finding the truth marks the success of the search for truth, the value of the latter is not limited to that of the former. Hence, cognitive activity is not just valuable because there is a chance that it might actually succeed; rather, the search for truth has value independently of its own success. Recall that I proposed that rational deliberation can be an epistemic good. More specifically, I would argue that it is epistemically good if it aims at truth, at getting things right.

This denies the epistemological thesis of instrumental veritism, which claims that all epistemic value reduces to either truth or to whether a state or a process contributes to finding the truth. (Sylvan, 2018) Coliva's notion of epistemic warrant that I considered earlier is instrumentally veritist: Only if there is some *truth indication* is a doxastic state epistemically warranted. I reject this instrumental veritism about epistemic warrant.

This raises the challenge<sup>24</sup> that cognitive activity is not an epistemic good, but epistemic on the one hand and good in a more general sense. Allan Hazlett (2016, p. 261) argues on the basis of this distinction that cognitive activity may contribute to most agents' well-being. Duncan Pritchard (2014b, p. 114) uses this distinction to defend instrumentalist veritism against different challenges. As Coliva's (2022, p. 175) challenge, that entitlement does not offer evidential support and is therefore not epistemically valuable, shows, I have to reject this distinction. Cognitive activity aimed at truth *is epistemically good*, but how?

Fortunately, Kurt Sylvan (2018, 2020) has developed a non-instrumentalist account of epistemic normativity that subscribes to veritism. His starting point is agreeing with the veritist that true belief (or accuracy) is the fundamental epistemic value. However, he diverges from the consequentialist veritist in arguing that there are not only instrumental derivative values but also non-consequentialist derivative values. Namely, Sylvan (2020, p. 9) argues that values also require *respect* – a state can also be *epistemically warranted* by manifesting respect for the truth (*cf.* Hurka, 2001). Cognitive activity aimed at truth essentially encodes respect for the truth; it consequently is warranted. Sylvan calls this view:

#### Value-First Epistemic Kantianism

- 1 **Valuing Thesis:** The fundamental normative explanation of why justified beliefs are justified is that they manifest certain ways of valuing fundamental epistemic value.

- 2 **Kantian Conception of Valuing:** The fundamental way of valuing epistemic value is to respect it.
- 3 **Veritist Conception of Fundamental Epistemic Value:** Accuracy is the fundamental epistemic value.

(Sylvan, 2020, p. 11)

Coliva and the sceptic overlook that we can be epistemically warranted by respecting the truth, not only by having evidence for it. While entitlements are grounded in a respect for truth, they do not only enable access to true beliefs but also other values. They ‘further the attainment of epistemic goods – of truth, understanding, and the anticipation of future experience, for example’ (Wright, 2014, p. 239). The question is whether these further epistemic goods, understanding or wisdom, are reducible to true beliefs. As I will argue in Chapter 4, some cognitive activity aims at such more complex goods (Zagzebski, 1996; Wright, 2010; Fleisher, 2017). If these epistemic goods can ultimately be reduced to a respect for the truth, I am happy to accept this. For simplicity’s sake, I will posit epistemic pluralism and treat them as independent epistemic goods for the rest of this work.

Does then *any* epistemic activity aimed at truth grant us entitlement? Even believing falsely or doing astrology? This problem lies at the heart of hinge epistemology and I will treat it in Chapter 3.

### 2.6.1 *Philosophy as a cognitive activity*

I have argued the search for truth or cognitive activity is epistemically valuable independently of its chance of success. As an illustration of this, consider philosophical research. I would argue that most philosophers consider what they do to be epistemically valuable. However, there is also a fairly widespread pessimism in philosophy as finding philosophical truths is very hard, almost impossible. This is illustrated by the sheer amount of philosophical disagreement (Chalmers, 2015). But this means that philosophical activity will have epistemic success on extremely rare occasions.

This leaves us with two options: Either philosophers are deeply irrational in pursuing philosophy and in considering their activity to be epistemically valuable, or their activity has an independent epistemic value. The first option does not appear palatable; therefore, I would argue that some or even many philosophers will be inclined to accept that cognitive activity has an independent value. Indeed, respect for the truth is built into philosophy’s name – love of wisdom.

Consider, for example, the Persian doctor and philosopher Ibn Sina. Clearly, his research was not a waste of time, it was epistemically valuable, although it was mostly false. But was it only of epistemic value because some centuries later, people started finding scientific truths thanks to his work? I doubt it (*cf.* Dellsén, Lawler and Norton, 2021).

Cognitive activity cannot begin from the empty set. You need a starting point: Cornerstones and rules. Consider that in classical logic too demonstrations from the empty set actually rely on a set of axioms – that is, rules that can be formulated as hinges. In other words, cognitive activity necessarily relies on a set of rule and cornerstone propositions.

Coliva and Doulas (2022, p. 251) argue similarly that there are *philosophical hinges* – the presuppositions that a philosopher needs in order to have philosophical reasons. Their starting point is also philosophical disagreement which they interpret as a special case of deep disagreement (Fogelin, 1985; Ranalli, 2020). However, they argue that not all philosophical disagreement is such irresolvable deep disagreement – which would be an untenable situation for philosophy. There are not only deep irresolvable *hinge philosophical disagreements* but also *intra-theoretical philosophical disagreements* (Coliva and Doulas, 2022, p. 12). These latter are philosophical disagreements among agents who share the relevant hinges. Consequently, they will be able to appeal to the same kinds of reasons and convince each other to resolve their disagreement.

Without trusting in some hinges, we would be unable to engage in philosophy and to search for truth. Hence, the only way to respect philosophical truth and realise the epistemic good of philosophical cognitive activity is by trusting in some philosophical hinges. This is the source of our entitlement to trust in the accuracy of our cornerstones and rules. We are entitled to trust in cornerstone and rule propositions because they enable us to respect and search the truth. This realises a non-consequentialist teleological value (Sylvan, 2020).

### 2.6.2 *Opting out*

We can also raise the Oblomovian challenge (Wright, 2014, p. 244; Coliva, 2015, p. 145) to this transcendental argument: Why search for the truth? Because we cannot help ourselves. We are biologically predisposed to engage with the world, as the example of core cognition shows.<sup>25</sup> It is part of what it means to be a human organism. If someone were to voluntarily spend her life in a sensory deprivation tank, not caring for any cognitive input or activity, we would deem this pathological. This naturalism tracks the transcendental argument developed by Moyal-Sharrock (2004) but anchors it to a greater degree in our biological and psychological make-up rather than in general anthropological considerations. My response to this question also distinguishes my proposal from Coliva's and Wright's: The answer of the latter has a strong consequentialist tendency and sees entitlement as serving the acquisition of knowledge. Coliva, meanwhile, takes epistemic rationality to be the source of its own warrant.

## 2.7 Ways to be entitled

In this chapter, I first examined whether hinges can be justified in any of the traditional ways. I argued that they cannot by examining whether the different ways in which we gain justification for our regular beliefs apply to hinges. But then this raises the question of whether we lack warrant for trusting in hinges.

I developed THE SIMPLE ARGUMENT that we are nevertheless warranted to trust in hinges. I have to note, however, that this simple argument begs the question against sceptics. The question is: How do we get this warrant, what is the source of the entitlement? Why is entitlement an epistemically good state?

I considered several proposals as to why entitlement is epistemically good. I introduced Wright's notion of entitlement in general as well as the sceptical arguments that motivate his project.

I then explicated in detail how, according to Wright, different kinds of entitlement can arise. First, I introduced strategic entitlement which arises from arguing that trusting in cornerstones is a dominant strategy. I rejected this consequentialist route as self-defeating. Second, I explained how entitlement of cognitive project works. This is the model for my own account of entitlement of cognitive activity. Third, I examined Wright's argument that our conceptual schema, our cognitive make-up, warrants entitlement. Fourth, I introduced the notion of entitlement of rational deliberation.

Next, I discussed another internalist account. Annalisa Coliva rejects the view that entitlement is possible because epistemic warrant requires the support of a proposition's truth. She instead argues that we are rational in trusting in cornerstones, although we are not warranted. Trusting in cornerstones constitutes rationality, therefore, it is *extendedly* rational to do so. I argued that Coliva's notion of epistemic warrant is too restricted.

I also presented Tyler Burge's account of externalist entitlement. I argued that there are interesting parallels between his and my account of entitlement, especially when we consider the internal perspectives in the two views and their biological background. Nevertheless, they do not fit very well, because Wright and Burge disagree on what justification and entitlement mean.

Finally, I presented my own view which takes entitlement of cognitive project as its starting point. I argued that we get entitlement in general from truth-directed cognitive activity. That is, I take investigation and cognitive activity to be epistemically valuable because they embody a respect for the truth. This is best illustrated by entitlement of cognitive project, but it extends to entitlement from our cognitive schema as well as to entitlement of practical deliberation – that is my account subsumes Wright's different proposals. I argued that these are not only consequentialistically

Table 2.3 Types of entitlements

	<i>Type of entitlement</i>	<i>Characteristics</i>
<b>Wrightean</b> (internalist)	Strategic entitlement	Consequentialist based on dominance reasoning
	Entitlement of cognitive project	Consequentialist or transcendental, based on the structure of investigation
	Entitlement of substance	Transcendental, based on the conceptual schema
	Entitlement of rational deliberation	Transcendental or pragmatist, based on practical rationality
	Extended rationality	Transcendental, replaces epistemic warrant with epistemic rationality
	Entitlement of cognitive activity	Transcendental, based on any truth-directed cognitive activity
<b>Burgean</b> (externalist)	Reason-independent warrant	Reliabilist warrant for perceptual beliefs formed in normal environments

valuable as a means to the end of getting at truth, as then we would fall into a similar problem as that which beset strategic entitlement. To put it simply, an epistemically active agent who never finds any truth but respects it is still better off than an epistemically passive agent because the former realises the good of epistemic activity aimed at truth. The former is transcendentially entitled to the hinges of her investigations.

## Notes

1. As I argued in the previous chapter, I take hinge certainties to simply stand on their own. They *de facto* do not rely on other hinges. If asked why we hold them, we could feel tempted to simply appeal to further hinges to justify them. This comes naturally, given that hinges do not occur as single atomic propositions, but rather as bundles that form an entire framework (Wittgenstein, 1969, §225).
2. Consider how easy it would be to fabricate evidence otherwise. Just take a piece of paper and write: ‘§1 Everything written on this paper is true, §2 . . . , §3 . . . , etc.’
3. This way of arguing goes back to the so-called Agrippan trilemma (Sextus Empiricus, 2000). I do not think that epistemic agents can be committed (implicitly) to an infinite regress of hinges.
4. Compare this to Quine’s (1951) attack on the analytic a priori where he argues that all conceptual content must be acquired at some point.
5. I use ‘feeling’ to point to the phenomenological quality of rightness that is ascribed to intuitions. These are also sometimes called ‘seemings’ (Conee and

- Feldman, 2008). It is the non-empirical counterpart to the phenomenology of perception (seeing, hearing, etc.), which arguably delivers justification for propositions a posteriori. These ‘feelings’ or seemings may also be interpreted as evidence.
6. There is also a newer epistemological project, which I will not discuss here. According to knowledge first epistemology (Williamson, 2001), we are justified in accepting a proposition if and only if we know it. Arguably that would also count for hinges. While highly successful as an epistemological project, I will not deal with this account given its diametrically opposed assumptions. It cannot account for the epistemological peculiarity of trust in hinges. We simply know or fail to know them – I do consider the thesis that we know hinges in Chapter 5.
  7. I believe that the generality problem (Conee and Feldman, 1998) confronts us with that very question: What unifies a faculty that is to be evaluated for reliability? I will return to this problem in Chapter 4.
  8. An alternative approach would be to argue that we gain justification for our hinges through inference from these reliabilistically justified non-hinge propositions. Through hearing, I am reliabilistically justified in my belief that the piano is out of tune. If I infer from this claim the hinge that my hearing works, is the hinge also reliabilistically justified? No, the hinge is what warranted this inference as an implicit premise. Hence, the reliabilist justification cannot transmit to the hinge. Recall the bootstrapping arguments from Chapter 1.
  9. The only way that this would seem unproblematic is if we postulate a 100% reliable faculty for hinges. This, however, goes against the naturalist underpinnings of reliabilism and I would be suspicious of any postulated faculty that is 100% reliable.
  10. Compare this to Coliva’s (2015, p. 59) arguments in the same direction.
  11. See also Coliva and Pritchard (2022, pp. 121–122). Crispin Wright (2014, pp. 219–220) emphasises that dogmatism will be unable to grant us a *rational claim* to justification for our basic beliefs even if you have the justification and that scepticism targets these claims. We will come back to this argument in Chapter 5.
  12. Except, maybe, Pyrrhonic sceptics.
  13. This brackets coherentism.
  14. Descartes’ (2013) own arguments were somewhat different. His epistemic standard was to reject any belief if he had the slightest reason to doubt it. In his posited sceptical scenario, he was created such that he systematically misjudged – rational insight was the epistemic gold standard for Descartes.
  15. See also (Coliva and Pritchard, 2022, pp. 17–27).
  16. In (Wright, 2004, p. 185), he worries whether this gamble can actually warrant something doxastically more robust than pragmatic assumption. By (Wright, 2014, p. 227), he argues that trusting our presuppositions nevertheless is the dominant strategy.
  17. With Cartwright (1983), you might argue that having false presuppositions does not necessarily lead to false beliefs. Note, however, that the false presuppositions that Cartwright uses are of a special kind, namely they track certain features of the proposition that is actually true. Hence, Cartwright’s false presuppositions are a special case and arguably in the minority among possible presuppositions.
  18. Though note that Pedersen (2020) advocates a view that attempts to broaden consequentialism by including other epistemic goods like coherence. Pedersen

- also argues that if the consequentialist warrant goes through, then why would we need entitlement? I am not too worried by this objection. Entitlement is simply the stand-in for non-achievement accounts of warrant, that is, the kind of warrant that consequentialist and transcendental arguments deliver.
19. Pedersen's (2009), Pritchard's (2014a), and Elstein and Jenkins' (2020) criticisms of consequentialist entitlement also apply to this argument.
  20. Note that even dreams are presented in this mode.
  21. This is related to the complaints by Pritchard (2014a, 2016, p. 80) and Elstein and Jenkins (2020). Wright (2014, pp. 238–239) responds by embracing a general, consequentialist notion of rationality that (strategic) entitlement supports. This general rationality includes the realisation of epistemic values among others.
  22. Burge argues that perceptual states are non-propositional. Hence, they cannot be reasons; only the fact that you are in such a propositional state can be cited as a reason and give justification over entitlement (Burge, 2003, p. 525).
  23. Mikkel Gerken (2020) has taken a different route, arguing that justification does not arise from a belief's being supported by reasons but from the *faculty* of reason, while entitlement is not produced by this faculty. Given that this proposal closely tracks a distinction that I will argue for in Chapters 4 and 5 between Type 1 (non-reason) and Type 2 (reason) cognition; I will come back to this later.
  24. I thank an anonymous referee for raising this worry.
  25. In Chapter 4, I will additionally develop an argument about what epistemic excellence means on the basis of our cognitive and psychological make-up.

## References

- Ayer, A. J. (1952) *Language Truth & Logic*. New York City: Dover Philosophical Classics.
- Bealer, G. (2000) 'A Theory of the A Priori', *Pacific Philosophical Quarterly*, 81(1), pp. 1–30. <https://doi.org/10.1111/1468-0114.00092>.
- Beaver, D., Geurts, B. and Denlinger, K. (2021) 'Presupposition', in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2021). Available at: <https://plato.stanford.edu/archives/spr2021/entries/presupposition/>.
- Bratman, M. E. (1992) 'Practical Reasoning and Acceptance in a Context', *Mind*, 101(401), pp. 1–16.
- Briesen, J. (2012) *Skeptische Paradoxa*. Paderborn: Mentis.
- Burge, T. (1998) 'Frege on Knowing the Foundation', *Mind*, 107(426), pp. 305–347.
- Burge, T. (2003) 'Perceptual Entitlement', *Philosophy and Phenomenological Research*, 67(3), pp. 503–548. <https://doi.org/10.1111/j.1933-1592.2003.tb00307.x>.
- Burge, T. (2020) 'Entitlement', in Graham, P. J. and Pedersen, N. J. L. L. (eds.) *Epistemic Entitlement*. Oxford: Oxford University Press, pp. 37–142. <https://doi.org/10.1093/oso/9780198713524.003.0002>.
- Carey, S. (2009) *The Origin of Concepts*. Oxford: Oxford University Press.
- Cartwright, N. (1983) *How the Laws of Physics Lie*. Oxford: Oxford University Press.
- Chalmers, D. J. (2015) 'Why Isn't There More Progress in Philosophy?', *Philosophy*, 90(1), pp. 3–31. <https://doi.org/10.1017/S0031819114000436>.



- Clifford, W. K. (1999) 'The Ethics of Belief', in Madigan, T. (ed.) *The Ethics of Belief and Other Essays*. Amherst: Prometheus, pp. 70–96.
- Coliva, A. (2015) *Extended Rationality*. Basingstoke: Palgrave Macmillan.
- Coliva, A. (2020) 'Against (Neo-Wittgensteinian) Entitlements', in Graham, P. J. and Pedersen, N. J. L. L. (eds.) *Epistemic Entitlement*. Oxford: Oxford University Press, pp. 327–343. <https://doi.org/10.1093/oso/9780198713524.003.0012>.
- Coliva, A. (2021) 'Strange Bedfellows: On Pritchard's Disjunctivist Hinge Epistemology', *Synthese*, 198(S15), pp. 3521–3532. <https://doi.org/10.1007/s11229-018-02046-z>.
- Coliva, A. (2022) 'Propositional and Doxastic Hinge Assumptions1', in Silva, P. Jr. and Oliveira, L. R. G. (eds.) *Propositional and Doxastic Justification*. New York: Routledge, pp. 162–180. <https://doi.org/10.4324/9781003008101-14>.
- Coliva, A. and Doulas, L. (2022) 'What Philosophical Disagreement and Philosophical Skepticism Hinge on', *Synthese*, 200(3), p. 251. <https://doi.org/10.1007/s11229-022-03735-6>.
- Coliva, A. and Pritchard, D. (2022) *Skepticism*. Oxon: Routledge. <https://doi.org/10.4324/9780429057946>.
- Conee, E. and Feldman, R. (1998) 'The Generality Problem for Reliabilism', *Philosophical Studies*, 89(1), pp. 1–29. <https://doi.org/10.1023/A:1004243308503>.
- Conee, E. and Feldman, R. (2008) 'Evidence', in Smith, Q. (ed.) *Epistemology*. Oxford: Oxford University Press, pp. 83–104. <https://doi.org/10.1093/acprof:oso/9780199264933.003.0005>.
- Davies, M. (2004) 'II – Martin Davies : Epistemic Entitlement, Warrant Transmission and Easy Knowledge', *Aristotelian Society Supplementary Volume*, 78(1), pp. 213–245. <https://doi.org/10.1111/j.0309-7013.2004.00122.x>.
- Dellsén, F., Lawler, I. and Norton, J. (2021) 'Thinking about Progress: From Science to Philosophy', *Noûs*, pp. 1–27. <https://doi.org/10.1111/nous.12383>.
- Descartes, R. (2013) *Meditations on First Philosophy*. Edited by J. Cottingham. Cambridge: Cambridge University Press.
- Elstein, D. Y. and Jenkins, C. S. I. (2020) 'The Truth Fairy and the Indirect Epistemic Consequentialist', in Pedersen, N. and Graham, P. J. (eds.) *Epistemic Entitlement*. Oxford: Oxford University Press, pp. 344–360. <https://doi.org/10.1093/oso/9780198713524.003.0013>.
- Evans, G. (1985) *Collected Papers*. Oxford: Oxford University Press. <https://doi.org/10.5860/choice.37-1485>.
- Fleisher, W. (2017) 'Virtuous Distinctions: New Distinctions for Reliabilism and Responsibilism', *Synthese*, 194(8), pp. 2973–3003. <https://doi.org/10.1007/s11229-016-1084-2>.
- Fogelin, R. (1985) 'The Logic of Deep Disagreements', *Informal Logic*, 7(1), pp. 1–8. <https://doi.org/10.22329/il.v7i1.2696>.
- Foley, R. (1991) 'Rationality, Belief and Commitment', *Synthese*, 89(3), pp. 365–392.
- Gerken, M. (2020) 'Epistemic Entitlement – Its Scope and Limits', in Graham, P. J. and Pedersen, N. J. L. L. (eds.) *Epistemic Entitlement*. Oxford: Oxford University Press, pp. 151–178. <https://doi.org/10.1093/oso/9780198713524.003.0004>.
- Graham, P. J. et al. (2020) 'Introduction and Overview', in Graham, P. J. and Pedersen, N. J. L. L. (eds.) *Epistemic Entitlement*. Oxford: Oxford University Press, pp. 1–34. <https://doi.org/10.1093/oso/9780198713524.003.0001>.

- Hazlett, A. (2006) 'How to Defeat Belief in the External World', *Pacific Philosophical Quarterly*, 87(2), pp. 198–212. <https://doi.org/10.1111/j.1468-0114.2006.00255.x>.
- Hazlett, A. (2016) 'Epistemic Goods', in Fletcher, G. (ed.) *The Routledge Handbook of Philosophy of Well-Being*. New York City, pp. 259–270. <https://doi.org/10.4324/9781315739793>.
- Hurka, T. (2001) *Virtue, Vice, and Value*. Oxford: Oxford University Press. <https://doi.org/10.1093/0195137167.001.0001>.
- James, W. (2014) 'The Will to Believe', in *The Will to Believe: And Other Essays in Popular Philosophy*. Cambridge: Cambridge University Press, pp. 1–31.
- Jenkin, Z. (2020) 'The Epistemic Role of Core Cognition', *Philosophical Review*, 129(2), pp. 251–298. <https://doi.org/10.1215/00318108-8012850>.
- Kant, I. (1998) *Critique of Pure Reason*. Edited by P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.
- Kolodny, N. (2005) 'Why Be Rational', *Mind*, 114(455), pp. 510–563. <https://doi.org/10.1093/mind/fzi223>.
- Lynch, M. P. and Silva, P. (2016) 'Why Worry About Epistemic Circularity?', *Journal of Philosophical Research*, 41, pp. 33–52. <https://doi.org/10.5840/jpr20165362>.
- McGlynn, A. (2014) 'On Epistemic Alchemy\*', in Dodd, D. and Zardini, E. (eds.) *Scepticism and Perceptual Justification*. Oxford: Oxford University Press, pp. 173–189. <https://doi.org/10.1093/acprof:oso/9780199658343.003.0009>.
- Moore, G. E. (1939) 'Proof of an External World', *Proceedings of the British Academy*, 25, pp. 273–300.
- Moyal-Sharrock, D. (2004) *Understanding Wittgenstein's on Certainty*. Basingstoke: Palgrave Macmillan.
- Pedersen, N. J. L. L. (2009) 'Entitlement, Value and Rationality', *Synthese*, 171(3), pp. 443–457. <https://doi.org/10.1007/s11229-008-9330-x>.
- Pedersen, N. J. L. L. (2020) 'Pluralist Consequentialist Anti-Scepticism', in Graham, P. J. and Pedersen, N. J. L. L. (eds.) *Epistemic Entitlement*. Oxford: Oxford University Press, pp. 297–326. <https://doi.org/10.1093/oso/9780198713524.003.0011>.
- Pritchard, D. (2014a) 'Entitlement and the Groundlessness of Our Believing\*', in Dodd, D. and Zardini, E. (eds.) *Scepticism and Perceptual Justification*. Oxford: Oxford University Press, pp. 190–212. <https://doi.org/10.1093/acprof:oso/9780199658343.003.0010>.
- Pritchard, D. (2014b) 'Truth as the Fundamental Epistemic Good', in Matheson, J. and Vitz, R. (eds.) *The Ethics of Belief*. Oxford: Oxford University Press, pp. 112–129. <https://doi.org/10.1093/acprof:oso/9780199686520.003.0007>.
- Pritchard, D. (2016) *Epistemic Angst*. Princeton: Princeton University Press.
- Pryor, J. (2000) 'The Skeptic and the Dogmatist', *Noûs*, 34(4), pp. 517–549. <https://doi.org/10.1111/0029-4624.00277>.
- Putnam, H. (1981) 'Reason, Truth, and History', *The Philosophical Quarterly*, p. 108. <https://doi.org/10.2307/2219218>.
- Quine, W. V. O. (1951) 'Two Dogmas of Empiricism', *The Philosophical Review*, 60, pp. 20–43.
- Ranalli, C. (2020) 'Deep Disagreement and Hinge Epistemology', *Synthese*, 197(11), pp. 4975–5007. <https://doi.org/10.1007/s11229-018-01956-2>.
- Reichenbach, H. (1938) *Experience and Prediction*. Chicago: University of Chicago Press.

- Rysiew, P. (2008) 'Rationality Disputes – Psychology and Epistemology', *Philosophy Compass*, 3(6), pp. 1153–1176. <https://doi.org/10.1111/j.1747-9991.2008.00178.x>.
- Sextus Empiricus. (2000) *Outlines of Scepticism*. Edited by J. Annas and J. Barnes. Cambridge: Cambridge University Press.
- Sosa, E. (1991) 'Intellectual Virtue in Perspective', *Knowledge in Perspective: Selected Essays in Epistemology*, pp. 270–293.
- Strawson, P. F. (1959) *Individuals*. London: Routledge.
- Sylvan, K. L. (2018) 'Veritism Unswamped', *Mind*, 127(506), pp. 381–435. <https://doi.org/10.1093/mind/fzw070>.
- Sylvan, K. L. (2020) 'An Epistemic Nonconsequentialism', *Philosophical Review*, 129(1), pp. 1–51. <https://doi.org/10.1215/00318108-7890455>.
- van Cleve, J. (2003) 'Is Knowledge Easy – or Impossible? Externalism as the Only Alternative to Skepticism', in Luper, S. (ed.) *The Sceptics*. Aldershot: Ashgate, pp. 45–59.
- Vogel, J. (2000) 'Reliabilism Leveled', *The Journal of Philosophy*, 97(11), pp. 602–623. <https://doi.org/10.2307/2678454>.
- Williamson, T. (2001) *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Wittgenstein, L. (1969) *Über Gewissheit – On Certainty*. Edited by G. E. M. Anscombe and G. H. von Wright. Oxford: Blackwell.
- Wright, C. (1985) 'Facts and Certainty', *Proceedings of the British Academy*, 71, pp. 429–472.
- Wright, C. (2004) 'I – Crispin Wright: Warrant for Nothing (and Foundations for Free)?', *Aristotelian Society Supplementary Volume*, 78(1), pp. 167–212. <https://doi.org/10.1111/j.0309-7013.2004.00121.x>.
- Wright, C. (2014) 'On Epistemic Entitlement (II)', in Dodd, D. and Zardini, E. (eds.) *Scepticism and Perceptual Justification*. Oxford: Oxford University Press, pp. 213–247. <https://doi.org/10.1093/acprof:oso/9780199658343.003.0011>.
- Wright, S. (2010) 'Internalist Virtues and Knowledge', *Acta Analytica*, 25(2), pp. 119–132. <https://doi.org/10.1007/s12136-009-0066-0>.
- Zagzebski, L. (1996) *Virtues of the Mind*. Cambridge: Cambridge University Press.

# 3 Problems for entitlement

## Demarcation, arbitrariness, and relativism

### 3.1 Introduction

In the preceding two chapters, I have argued that we are entitled to trust that hinge propositions are true. We are entitled and thereby warranted to do so because hinge propositions are necessary preconditions that transcendently enable us to be cognitively active and to search for truth.<sup>1</sup>

You might ask: What more do we want? We have found an internalist warrant to trust in fundamental hinges. Truth-respecting cognitive activity is a positive condition for entitlement in the sense that it is a source of epistemic warrant. Hence, cognitive activity generates entitlement. You get warrant for nothing, and the hinges for free!

Unfortunately, this is not the end of the story. I mentioned that there are also negative conditions on entitlement. Negative conditions require the absence of properties that would defeat entitlement. Wright proposes the following negative condition as necessary for entitlement: There must be ‘no sufficient reason to believe that [a hinge] is untrue’ (Wright, 2004, p. 183). This negative condition guarantees that we are *rational* when we are entitled to trust in a hinge. In this chapter, I will argue that we need more negative conditions than just a no-defeater condition.

This is especially the case given that, on my view, this no-defeater condition does not work for entitlement. Wright thinks that entitlement is asymmetrically sensitive to evidence: There can be no non-circular evidence for entitlements, but there can be counterevidence. Given the peculiar role of hinges, I would argue that entitlement is symmetrically insensitive to evidence: There can be neither a non-circular support for, nor a defeat of, hinges and hence entitlement. That means, on my view, we do not get any mileage out of a no-defeater condition.

Negative conditions serve the function of distinguishing real entitlements from merely apparent and flawed entitlements. Beyond preserving rationality, negative conditions must stave off a different threat: Relativism and the problem of demarcation.

In his papers on epistemic entitlement, Crispin Wright (2004, 2014) mentions that the *problem of demarcation* arises for entitlement. The problem of demarcation originated in philosophy of science, notably with Karl Popper (1962). It asks: What is the criterion that demarcates a good scientific theory from a bad pseudoscientific theory? Similarly, we may ask: What demarcates a genuine entitlement from a merely apparent entitlement?

The first question here is: What does the problem consist of? The second question is whether the problem of demarcation really arises for entitlement, that is, whether it is a real problem for entitlement. I will therefore first present Crispin Wright's brief considerations on this issue. On the one hand, Wright adduces a range of bizarre beliefs that are candidate beliefs for entitlements. Their bizarreness appears to be incompatible with entitlement. On the other hand, Wright thinks that trust in cornerstones does not appear to be a case of 'responsible belief management' (Wright, 2014, p. 245), thus threatening their entitlement status.

I will argue that bizarreness gives no principled grounds to deny entitlement for an instance of epistemic trust. It is, however, sufficient to make the problem of demarcation salient and pressing: *That* kind of belief should not be warranted, and therefore there should be no entitlement for it.

This leads to a further important observation: In its current form, absent any negative conditions, my account of epistemic entitlement opens the gates to a radical epistemic relativism. If you are committed to search for truth in a certain way, then you are entitled to trust in the pertinent hinges. Entitlement of cognitive activity imposes no methodological constraints whatsoever and without any negative conditions you can gain the correspondent entitlements for just about any rule or cornerstone proposition.

This observation tracks an important current debate in non-evidential epistemology (Coliva and Moyal-Sharrock, 2016). A prominent position in hinge epistemology, exemplified by Moyal-Sharrock (2004) argues that hinges cannot be part of our epistemology, because hinges are rules and not truth-apt. Consequently, they cannot be warranted. Duncan Pritchard alternatively argues that given the pre-evidential function of hinges, we cannot rationally come to believe a hinge, thus these cannot be warranted either (Pritchard, 2016, p. 98). Consequently, hinges too are beyond warrant and entitlement. This obviously is not the view of hinges that I defended in preceding chapters.<sup>2</sup>

Relativism is also implied by a form of justification I discussed previously: Coherentism. Two equally coherent but incompatible belief systems will be equally justified and therefore on equal standing – because justification is a property internal to a belief system,<sup>3</sup> according to coherentism.

I want to argue that this unrestricted relativism is ultimately incompatible with entitlement. If a belief is one of two equally reasonable alternatives, then the belief is epistemically *arbitrary*. I will further argue that

epistemic arbitrariness threatens entitlement. It may not directly undermine or defeat an entitlement, but it will threaten your higher order claim that you are entitled to trust this arbitrary hinge.

Jochen Briesen (2012) has developed a strong and rich account of the negative and positive conditions required for entitlement of cognitive project. This account also deals with relativism and bizarre hinges; I will therefore draw on it to illustrate the problem of demarcation. I will argue that the conditions laid down by Briesen only preserve an agent's rationality with entitlement. They fail to demarcate bizarre beliefs and they only partially succeed in dealing with relativism. I will also examine how Jochen Briesen's proposed solution to the problem of demarcation fares with respect to my diagnosis of arbitrariness. I will argue that although the conditions he proposes do guarantee the rationality of someone entitled to trust, they cannot distinguish flawed entitlements from the real article.

The goal of this chapter is to establish that the positive conditions I proposed – namely, that we gain entitlement from cognitive activity that aims at truth – are not sufficient to give us a claim to entitlement on their own. We also need to satisfy negative conditions, namely, entitlement needs to avoid arbitrariness, otherwise entitlement would become relativist and our claim to it would be defeated by its arbitrariness.

### 3.2 The problem of demarcation

According to Wright, entitlement faces an unresolved problem: How do we distinguish genuine entitlement from cases that seem to fulfil the criteria of entitlement but which should not constitute instances of epistemic warrant?

The point has not gone away that it is not in general, or even usually, consistent with responsible belief management to accept things without evidence or relevant cognitive achievement. What are the principles that determine when one may do so and when one ought not? How do we distinguish the genuine entitlements from the prejudices, mere assumptions and *idées fixes*?

(Wright, 2014, p. 245)

That is, Wright recognises that there need to be some constraints – negative conditions – on what counts as an entitlement in addition to the positive conditions laid out in the preceding chapter. In the case of entitlement of cognitive project, most plausibly, this boils down to restrictions on what sort of investigation or cognitive project generates its own entitlement. Analogously, on my proposed view, we may ask what negative conditions cognitive activity needs to fulfil in order to generate epistemic warrant for its hinges.

Note that I already built a condition that could be considered negative into my proposed account of entitlement of cognitive activity. I did not extend entitlement to just any kind of cognitive activity. I restricted it to cognitive activity that *aims at finding the truth*. This restriction excludes cases of self-deception, such as a bank robber fabricating cover stories from generating entitlement for her cognitive activity, even if she ends up bringing herself to believe those stories.<sup>4</sup> However, this constraint is not enough to solve the problem of demarcation.

You might wonder why demarcation is a problem at all. Maybe generosity is one of the quirks of entitlement: All sorts of cognitive activity could generate entitlement.<sup>5</sup> Another passage by Wright vividly illustrates the difficulty with this suggestion:

What are the barriers to an entitlement to wood spirits, ectoplasm, gods, and a plethora of existing but non-actual spatio-temporally unrelated concrete possible worlds?

(Wright, 2004, p. 204)<sup>6</sup>

If you were committed to the cognitive project of figuring out your late great-grandmother's unfulfilled wishes and this project happened to presuppose that there is a thing called ectoplasm, then you would be entitled to trust in the existence of ectoplasm. You've got to start somewhere, why not at ectoplasm? So now you're warranted to trust that there is ectoplasm. I find that objectionable, I hope you do too. An epistemology that generates warrant for the existence of ectoplasm is too generous. If my account allows for entitlement to such bizarre entities, then something is wrong with my account.

Another kind of similarly questionable cognitive activity is sceptical investigation. Namely someone may, for whatever reason, end up presupposing that they are a brain in a vat or part of a simulation. They may then undertake investigations to understand the nature of their predicament, find out the rules of this simulation, etc. Given that this is a cognitive activity aimed at truth and no non-circular reason speaks either for or against such scepticism, they are then entitled to trust in the hinge that they are a victim of a hallucination or simulation.

I think that these examples show that we need to pay attention to the problem of demarcation. But, as I will argue here, the bizarreness of these examples is not the root of the problem. Rather it is a symptom that in and of itself gives us enough reason to try to solve the underlying problem.

So, what exactly is the problem with bizarre certainties like 'wood spirits, ectoplasm, gods,' etc.? The simple answer seems to be: They are obviously false. There is clearly no such thing. Obvious falsehood should clearly be a barrier to and incompatible with epistemic warrant and hence

entitlement. Pointing to obviously false hinges would seem to be a straightforward answer to the problem of demarcation because it proposes a clear negative condition.

Unfortunately, we cannot congratulate ourselves for having solved the problem of demarcation. Bizarreness is only one of the unfortunate consequences of the problem, but it is not the problem itself. To see this, consider the following two reflections.

First, not everything that appears bizarre is actually false; *a fortiori* it is not obviously false. There are seemingly bizarre phenomena in this world of ours: Fish that change their sex several times in the course of their life, black holes at whose event horizon time appears to freeze and space gets stretched, and people giving away goods and services on the basis of a signature. Somebody who has never heard of these things might consider them to be obviously false, however, they do occur.

Second, there are obvious falsehoods that are not bizarre. Consider mistaken mathematical or mechanical beliefs, for example, the idea that heavy objects fall faster than lighter ones. This is obviously false, but it would not be bizarre if someone were to believe it.

What this shows is that bizarreness itself does not really track obvious falsehood. Additionally, I believe that bizarreness tends to be simply a product of your hinges – bizarreness is world-view relative. A conviction is not called bizarre because it is obviously false, but because we cannot understand how someone would come to believe something like *this*. It is bizarre to imagine oneself believing the same. Bizarreness means that one is unable to put oneself into that position because one lacks the relevant frame of reference, that is, hinges.

What makes a belief seem bizarre to you is often a lack of the relevant hinges that provide a doxastic route to acquiring this belief. Given that hinges are usually implicit, we tacitly assume they are shared. If someone utters a belief based on divergent hinges, then this belief will not fit with your world view. The belief will appear bizarre.

As an illustrative example consider the tribe member that Wittgenstein (1969, §106) mentions who tells you that they were on the moon last night. You do not know what to make of this, it is obviously false. Going to the moon is not something you do just like that. The idea is utterly bizarre. The person clearly has hinge certainties that diverge from your own.

But if bizarreness itself is a mere product of divergent hinges and we are entitled to these certainties, then bizarreness cannot be a criterion for adjudicating whether we are in fact entitled to trust in these hinges. Bizarreness only tells us how well some candidate proposition for belief fits with our own world view. Bizarreness can therefore illustrate why we feel that an unrestricted entitlement is problematic, but it is no solution to the problem of demarcation.



### 3.3 Relativism

I believe that in the background to Wright's worry about bizarre beliefs lurks a further and more general problem. If there are no constraints on what can count as an entitlement, it is not only bizarre and troubling hinges that can claim the mantle of epistemic entitlement, but *just about anything can*. An indiscriminate mass of world views would be on equal standing – entitlement without demarcation invites a pernicious relativism.

Paul Boghossian (2006) defines relativism in its most general sense as a claim that certain propositions are relative to some frame of reference, which may vary. Given the semantic function of rules and cornerstones, any belief system will be relative to a set of hinges. Just as there are many ways to skin a cat, there are many ways to understand the world.

One way to characterise epistemic relativism is the thesis that there could be many equally warranted but mutually incompatible convictions or views of the world. In terms of entitlement, this would mean that Aisha, Bertie, and Carla each can have fundamentally different hinges – assume Aisha to be a scientific realist, Bertie a theistic realist, and Carla a scientific anti-realist. Yet they all are equally entitled to trust in their fundamental hinges. They all are equally epistemically warranted concerning their fundamental certainties. Namely, they are entitled and there could even be further possible world views that are equally warranted.

While relativism is something of a bogeyman in philosophy and I would follow a time-honoured tradition by simply declaring it to be bad, I want to go deeper into why relativist entitlement is problematic. One important point to note is that, on my view, competing alternative world views all turn out to be evidentially adequate. This is because entitlement arises out of hinges and hinges determine relations of evidential support. Thus, different entitlements may generate internally equally coherent and evidentially supported belief systems. (Hazlett, 2014)

In the broad sense, entitlement is clearly relativist: Evidential warrant is *relative* to entitlement. I also believe that this phenomenon will never really disappear from an account of entitlement. It is not the case that there is only one unique set of entitlements while all other belief sets lack entitlement. If this were so, would not simply the unique true belief set be entitled? Entitlement would be a true-belief externalism. If our theory of entitlement did not allow several alternatives, then why bother with such a theory? Why not just claim that the one true theory is warranted by virtue of its truth and coherence?<sup>7</sup>

But the concern about relativism here is not with the idea of relativism itself. Rather, it is about an unbounded and unrestricted relativism that swamps us with infinitudes of self-warranting theories. This is also something that card-carrying relativists would not support (Kusch, 2019, p. 285).

Hence, in order to avoid falling victim to a pernicious relativism of infinitely many self-warranting theories, entitlement requires negative conditions

that block hinges from gaining warrant gratuitously, that is, on the basis of *purely formal or structural* properties. If warrant depends solely on the internal structure of a belief set, that is, whether this is consistent and coherent, then this insulates the belief set in an important way. There may be arbitrarily many different incompatible possible belief sets with exactly the same structural properties – and these world views would be equally warranted relative to themselves.

Consequentialist entitlement, for example (Pedersen, 2020), is also open to this relativism. This is because in the sphere of fundamental hinges any hinge that promises potentially good epistemic outcomes is fair game. As soon as the hinges fulfil the minimal structural requirements, we may bet our entitlement on their being right.

Similarly, coherentism is also a form of relativist warrant. It is the thesis that epistemic warrant arises from how our beliefs cohere with each other; that is, from how integrated our theory is. Coherentism generates warrant purely from out of formal or structural properties. Thus, it also opens up an avenue for relativism: All coherent theories are equally warranted, because every theory is only justified through coherence relative to itself. The coherentist is stuck in this situation; any two equally coherent theories will be equally justified.

Annalisa Coliva also raises the issue of relativism for her account (Coliva, 2015, p. 140). However, she is only worried about alternative basic epistemic practices, that is, first-order rules about our perception and basic inference. In other words, according to Coliva, the only threatening relativism concerns our common sense faculties, where we might have idealist or phenomenalist alternatives. There is an alternative possible rule that our experience does not tell us about objects, but is simply about appearances and sense data. Our common sense rule therefore is arbitrary.

Coliva rejects this by arguing like Reid (2012) that *de facto* these alternative basic rules are not available. We can therefore disregard this option. While my own argument goes in a similar naturalist direction, I do not think that it is sufficient to point to our basic capacities. As one point in case, consider the fact that some people claim additional basic faculties and the corresponding hinges for themselves: From extraordinary mathematical intuition to clairvoyance or a *sensus divinitatis*, there have been claims made to many basic faculties, some more and some less controversial. I will come back to Coliva's argument in Chapter 5.

### 3.4 Arbitrariness

I believe that there is a further, even deeper problem with unrestricted entitlement. My thesis is that on the accounts presented up to now, epistemic entitlement is *arbitrary*, because these accounts mostly present positive conditions for generating warrant, while the negative conditions restricting

it are too weak. This includes my own account of entitlement of cognitive activity in its current version. Arbitrariness arises if it remains underdetermined which of at least two options should be picked. I argue that arbitrariness threatens epistemic warrant.

I define arbitrariness as follows:

- (A) A belief that P is *arbitrary* if and only if there is no criterion independent from whether P that makes P preferable to some incompatible alternative Q.<sup>8</sup>

Arbitrariness applies to hinges because they are essentially pre-evidential. They are not accepted on the basis of any evidence and absence of evidence implies that there is nothing speaking either for or against their truth. Entitlement on the accounts I proposed in the preceding chapter does not bring forth anything that speaks in favour of or against particular hinges either. Instead, it points to what we gain by trusting hinges. Nothing favours one hinge over an incompatible alternative hinge. The two hinges will simply enable different investigations. But which investigation or type of cognitive activity I pursue is arbitrary.

So if entitlements are arbitrary, why would arbitrariness threaten entitlement? Why is arbitrariness a problem for epistemic warrant? With regular beliefs, arbitrariness means the absence of evidence favouring it over incompatible alternatives. This implies a missing evidential justification. The principle at play here is the so-called underdetermination principle: 'If q is a competitor to p, then one can know p only if one can *non-arbitrarily* reject q.' (Vogel, 2004, p. 427, my emphasis). Brueckner formulates the same principle for justification (Brueckner, 2005, p. 388).

But we should not extend this consideration, that is, that arbitrariness defeats justification, willy-nilly to entitlement. After all, entitlement is supposed to work without evidential support. I will argue that arbitrariness is a problem for warrant because there is also a kind of non-evidential underdetermination principle.

We can consider the issue again with the example of the road to Larisa. If I lack any information on how to get to Larisa, I make an arbitrary choice between the two possible roads. Obviously, it is pragmatically reasonable to arbitrarily pick one of the two roads if I need to get to Larisa. But there seems to be no epistemic warrant for either option to be had. However, with Larisa, there is evidence to be found in principle; after all, it is at a determinate distance from where you are and people have been there. That means, it is only *contingently arbitrary* which road I pick. Whether my choice is indeed the road to Larisa can in principle be vindicated through evidence.

But what would it mean to vindicate a hinge that precedes and underwrites all subsequent investigations? Vindicating hinges in competition with incompatible alternatives is not possible except either by relying on the hinge itself, which would be circular, or by introducing a new hinge. Therefore, entitlement would be arbitrary as it cannot be vindicated without circularity or regress. Note that we can circularly support any hinge with itself; any hinge is equally self-warranted. For that reason, hinges and entitlement are *necessarily evidentially arbitrary*.

Note that the arbitrariness of hinges extends beyond their evidential status: We gain entitlement through *the same type* of argument each time. This positive condition is applicable to plenty of incompatible alternative hinges. Entitlements are thus not only evidentially but also generally arbitrary. There are no epistemic reasons distinguishing any option making it preferable to its competitors, at least if we do not strengthen the negative conditions such that they differentiate between different alternatives.

Arbitrariness is a problem for epistemic warrant because it rationally underdetermines the choice of our hinges. We gain entitlement from cognitive activity that is *aimed at truth*. If it is rationally underdetermined which of several incompatible hinges I should pick, then this truth aim gets hampered. I am not aiming at truth when I haphazardly pick one hinge and type of cognitive activity over others. I could as well pick *just about any* other type of activity. This also leads to a relativist explosion. If entitlement has no restraining or discriminatory negative conditions, it generates plenty of alternatives that all have an equal epistemic standing. Entitlement therefore needs further barriers to reduce the number of available alternatives.

Entitlement is warrant to trust something to be true in the absence of evidence. But are we warranted in trusting in any arbitrary hinge? The appropriate doxastic state for an epistemically arbitrary situation is not trust, but agnosticism (Friedman, 2013). This arbitrariness may not even threaten entitlement of cognitive activity *per se*; Pedersen (2022) argues that generosity and therefore a certain degree of arbitrariness is part and parcel of the entitlement view. But even if arbitrariness does not *directly* undermine and defeat entitlement, it will still undermine one's *claim to entitlement* (Wright, 2004, p. 208), that is the ability to explicitly appeal to this warrant. For an internalist, the loss of the higher order warrant to claim entitlement may prove similarly devastating and cannot be bridged by externalism.<sup>9</sup>

That is, to get from warranted agnosticism to warranted trust, that is, entitlement or a claim to entitlement, we need to create *some* relation to truth or exclude alternative objectionable hinges.<sup>10</sup> I suspect that arbitrariness is a gradual affair, a hinge may be more or less arbitrary; for example, default entitlement to any proposition would be maximally arbitrary.

Without the constraints of some negative conditions, candidate entitlements are too arbitrary to be claimed.

In sum, arbitrariness underlies the problem of demarcation. Arbitrariness is not ‘consistent with responsible belief management’ (Wright, 2014, p. 245) because it means being detached from the goals of epistemic activity. For these reasons, arbitrariness and the problem of demarcation need to be reduced, as well as the threat of pernicious relativism that arbitrariness entails. Namely, entitlement needs more stringent negative conditions than I have proposed up to now. These negative conditions will exclude some hinges, which would reduce the arbitrariness of the remaining options.

The same issue arises for coherentist warrant, given that it is a form of relativism. Coherentist warrant only works on the basis of an already given theory, it only helps you to rationally expand such a theory. Its starting point or basis is nonetheless arbitrary.

There is a way to make a coherentist warrant track the goal of epistemic activity: Making truth itself coherentist. That is, a statement is true in coherentism if and only if it coheres with other true statements (Young, 2018). In such a case, coherentist justification would closely track coherentist truth, because a principal way in which coherence between propositions is generated is through entailment or epistemic support.

If our account of entitlement only worked for coherent truth, the project would be very weak. Entitlement should also work for more demanding notions of truth that are widely accepted. I take at least some instances of truth to consist in correspondence; notably, everyday assertions about medium-sized dry goods are something that coherentist truth does not capture well. Coherentist truth is also arbitrary, just like the coherentist justification tracking it because it collapses epistemology and semantics into one.

A different problem that has been treated more extensively in the debate on entitlement is the so-called ‘leaching problem’ which is credited to Sebastiano Moruzzi (Wright, 2004, pp. 208–209, 2014, p. 228; McGlynn, 2017). The problem is that in trusting in a hinge, we take a risk. There is no evidential or reliabilist vindication of the corresponding entitlement. But if the presuppositions for our cognitive activity in general involved risk-taking, then the whole enterprise is risky. Any regular belief that relies on a risky entitlement will inherit the latter’s risk. Therefore, risk also leaches into our evidential beliefs.

Arguably, the arbitrariness of entitlement is closely related to its riskiness: It seems that any arbitrary decision involves a risk. Arbitrariness will leach upwards just like risk. If the entitlements on which my beliefs rely are arbitrary, then these beliefs will inherit this arbitrariness. The arbitrariness of our world view will then spread to our particular beliefs, making them arbitrary. Even if entitlement is not defeated by arbitrariness, the fact that we *lack a claim* to arbitrary entitlements will undermine the internalist justification of our ordinary beliefs. If your ordinary belief presupposes a hinge proposition

for which you cannot claim entitlement, then how could you claim justification for it? Many kinds of internalist justification require that you be able to claim them, that is, have some sort of epistemic access to them.

All of these considerations show that we need to reduce the arbitrariness of entitlement because it is an epistemic flaw at the basis of our belief systems. Doing so will also deal with the threat of an unbridled relativism. If we reduce the arbitrariness of entitlements, then we also reduce the range of the relativist options.

### 3.5 Conditions for entitlement on Briesen's account

As mentioned earlier, Jochen Briesen's (2012) model of entitlement for a cognitive project is arguably the most elaborate account of this view. Unfortunately, it has received little attention in the debate, arguably because it has only been published in German. Briesen develops an account of entitlement to trust the presuppositions of questions we are committed to answering – answering a question is a cognitive project. I will illustrate how the problem of demarcation in terms of arbitrariness gets a hold on Briesen's account of entitlement of cognitive project.

In his book, Briesen proposes jointly necessary and sufficient conditions for entitlement of cognitive project, which he also takes to be capable of dealing with the problems of relativism and bizarre presuppositions. That is, Briesen thinks that his proposal solves the problem of demarcation. His characterisation goes as follows:

S's belief that  $p$  is warranted independently of truth-indicating factors [*i.e.*, an *entitlement*] at  $t_0$  if and only if

- (i)  $p$  is the presupposition of a rational and promising cognitive project  $P_0$  in which  $S$  is interested at  $t_0$ ;
- (ii) there is no other rational and promising cognitive project  $P_1$  that determined and made accessible a response (some time) before  $t_0$ , which implies that  $\neg p$  or implies that the process that led to the belief that  $p$  was unreliable;
- (iii)  $S$  has no available defeating reasons against  $p$ ;
- (iv) the inferential justification for  $p$  is circular in a certain sense.

(Briesen, 2012, p. 251, *my translation*)

This definition has a complement: A cognitive project is only *rational* if its presuppositions are consistent. A cognitive project is only *promising* if its presuppositions entail that the cognitive project can establish at least a partial response to the investigated question (Briesen, 2012, pp. 248, 250).

Briesen does not explicitly mention the problem of demarcation. However, he takes condition (i) to prevent the issue of bizarre or, in his words, ‘abstruse or daft projects’ (Briesen, 2012, p. 251) and condition (ii) to prevent a relativist explosion of arbitrarily many incompatible cognitive projects. (Briesen, 2012, p. 252) That is, he does not equate the problem of relativism with that of abstruse projects.

Briesen provides no deeper analysis of what exactly the issues addressed by (i) and (ii) consist of. He thinks that Wright’s account of entitlement of cognitive project has a problem with regard to abstruse and relativist entitlements and that this can be resolved by limiting the cognitive projects that can be a potential source of entitlement.

Given his conditions, Briesen seems to consider abstruse or bizarre cognitive projects to be simply irrational. Namely, abstruse or daft projects can be prevented by excluding projects with inconsistent presuppositions or presuppositions that do not actually permit any investigation – that is, that are doxastically or strategically irrational. This is a very simple solution.

Further, Briesen does not explore in great detail what might be problematic about relativism. Instead, he takes the problem to be illustrated by the possibility that there could be astrological cognitive projects. That is, he appeals to the consideration that if our theory generates a warrant for astrology, then something must have gone wrong with our epistemological theory.

His solution to the problem of a relativist explosion is to restrict the available presuppositions to those that have not yet been refuted. If preceding cognitive projects have already established that the stars do not influence your romantic prospects, then we cannot claim entitlement for this presupposition. This solution to relativism has an interesting feature: We cannot get entitlement for astrology today because we have carried out cognitive projects that show it does not work. But in ancient Mesopotamia, this was not the case, thus the Mesopotamians were entitled to claim warrant for their astrological presuppositions.

I would argue that the conditions proposed by Briesen are at least implicit in Wright’s own characterisation of entitlement. Condition (i) simply prevents irrationality. I do not think that Wright even considered it worth mentioning that we cannot have irrational, that is, inconsistent, cognitive project.

Similarly, I believe that (ii) is simply an instance of (iii). Wright does require something in this vein when he demands that ‘we have no sufficient reason to believe that [the presupposition] P is untrue’ (Wright, 2004, p. 191). Clearly, the results of previous cognitive projects provide such reasons. Further, condition (iv) designates the presupposition function necessary for entitlement which also Wright requires. Note that this requirement partially coincides with the requirement that entitlement applies only to hinges which always play this presupposition role.

Consider also that Wright seemed to think that the problem of demarcation arises even when we rule out irrationality and counterevidence. After all, this remains an outstanding issue at the end of his (2014) paper.

The obvious problem with condition (i) as regards the problem of demarcation is that nothing appears to conceptually exclude coherent and consistent but abstruse cognitive projects. Just like bizarreness, abstruseness does not imply inconsistency or evident falsehood. If it did, establishing this would require a lot of work. The condition is thus apparently too weak to deal with bizarre but consistent cognitive projects, which form an instance of the problem of demarcation.

I am also not sure about the relativism-blocking merits of condition (ii). If we assume that there are infinitely many relativist alternative presuppositions available and consider that past cognitive projects are, by the nature of history, limited in number, then ruling out relativist alternatives will mean playing a never-ending game of catch-up. That means condition (ii) does not give a principled limitation on the relativist availability of alternative presuppositions. In this sense, it is too weak.

Additionally, I suspect that condition (ii) might get us stuck on the wrong track. Nothing in the concept of a finished cognitive project guarantees its infallibility. But if a past cognitive project  $P_1$  got it wrong and mistakenly appeared to show that the true presupposition  $p$  is false, then we will never be able to gain a warrant for the cognitive projects that would require  $p$  as a presupposition, at least as long as project  $P_1$  is accessible.

It may very well be that the only way of debunking  $P_1$  would be through cognitive projects that require presupposition  $p$ . At the least, rejecting old apparently successful cognitive projects on the basis of new cognitive projects whose presuppositions are incompatible with the results of the old project seems to happen often. This seems to be the mechanism behind 'paradigm shifts' as Kuhn (1996) defends them. General relativity theory, for example, had to do away with many of the established results of classical mechanics in order to even be formulated.

I am also not sure whether condition (ii) can gain any traction on hinges. By their nature, hinges precede evidential support *and defeat*. If they were open to cognitive projects that could defeat them, then they would not be hinges as there would be a way of establishing that *this* is not the right hinge. Thus (ii) overlooks the fundamentality of convictions requiring entitlement.

If my diagnosis concerning bizarre hinges and relativism is correct, then the two problems result from the hinges' arbitrariness. Briesen might want to contest this diagnosis, but nevertheless the problem of demarcation in terms of arbitrariness remains. I will therefore examine whether or how his two conditions fare with respect to the problem of arbitrariness.



Clearly, the requirement of rationality posited by (i) reduces the arbitrariness of an entitlement or cognitive project, but it does not do so sufficiently. Rationality is the bare minimum required for entitlement but there is still plenty of space for arbitrary but consistent and coherent hinge beliefs. Thus, (i) is unmistakably a necessary condition for entitlement, but it does not sufficiently address the problem of demarcation.

(ii) also reduces arbitrariness, but I doubt whether it does so in a way that supports conferring an epistemic warrant. It tries to recruit our already established knowledge to avoid arbitrariness. However, in my eyes, this ties the subject too strongly to established presuppositions and hinges. It seems eminently reasonable to follow your contemporaries' world view if you want to make epistemic progress, because epistemic progress becomes much easier when relying on others and with the possibility of engaging in an epistemic exchange with them. But I doubt that this pragmatist point plays a role in acquiring an epistemic warrant. That is, I do not think that relying on your contemporaries or taking into account their research is either necessary or sufficient for entitlement.

It rather seems that certain epistemically revolutionary projects must do away with previously established results, because the latter could have rested on mistaken presuppositions. Consider, for example, the ever-changing presuppositions about what a polyhedron is in Lakatos's *Proofs and Refutations* (1976). These revolutionary approaches arguably also start off with entitlements for their hinges. Briesen's condition (ii) would thus tend to make entitlement much too conservative. It therefore cannot be a necessary condition for entitlement.

For all these reasons, I do not think that Briesen's negative conditions are able to resolve the problem of demarcation. Requiring rationality to avoid bizarreness does too little and the threat of relativism cannot be fended off by simply pointing to past research either because it overshoots by excluding worthwhile cognitive projects too and thereby potentially misleads us.

### 3.6 Demarcating arbitrary entitlements

The problem of demarcation is clearly baked into the foundations of the concept of epistemic entitlement. This arises from the fact that entitlement needs to function without any sort of epistemic favouring; there is no evidential or reliabilist justification to be had for our hinges (Coliva, 2015, p. 64). This absence of a direct supporting connection for truth needs to be mitigated if entitlement is to be a form of epistemic warrant.

The missing truth connection manifests itself in the epistemic arbitrariness of hinges and entitlement. This arbitrariness permits the well-known phenomena of bizarre or absurd beliefs that follow from the acceptance of cornerstones that we do not share. Unbounded relativism also results from

arbitrariness. Epistemic arbitrariness arises because hinges are epistemically unconstrained while the traditional constraints of evidence or reliability are not available here. Coherentist justification for hinges is also arbitrary, given that this merely constrains other beliefs but not the whole belief set. While arbitrariness needn't directly undermine the entitlement itself, it defeats one's claim to entitlement. Through leaching, this defeated claim to entitlement will spread to defeat your claim to ordinary justification.

Currently, on my view of entitlement, the agent's motivation to find the truth that guides the entitled cognitive activity is the only connection to truth. Cognitive activity must be guided by the goal of finding truth to generate entitlement. This is a very weak connection to truth and it leads to the problem of demarcation, given that one can aim at truth with arbitrary hinges. Nikolaj Pedersen (2009, p. 455) calls this feature 'strong realisation independence.' On this account, entitlement does not even require the possibility of realising true beliefs. The motivation to get there suffices.

Note that a similar problem also occurs within ethics. The bare will to do good or to do the right thing is not sufficient for moral agency. Any immoral action may appear to be justified or right to someone who is motivated to do good if they have the wrong background beliefs.

We need a stronger connection to truth than a mere motivation to get entitlement off the ground. Similarly, ethicists require a stronger connection to the moral good than a mere good will. Arguably, the oldest solution to this problem in ethics is to require virtue from the epistemic agent (Montmarquet, 1992). Virtue here means character dispositions that manifest and anchor the good will and that lead the agent to right actions. Note that Wright (2004, pp. 210–211) also draws a comparison between epistemology and ethics and requires 'intellectual integrity.' One way of cashing out integrity is to require Montmarquet's (1992) virtue of conscientiousness.

I will argue that epistemic virtue can solve the problem of demarcation for non-evidential warrant. Namely, for entitlement to avoid arbitrariness, it needs to arise from epistemically virtuous cognitive activity. Epistemic virtue will strengthen truth-directedness over mere motivation, without requiring evidential support. But for this project to succeed, I will first have to explain what I understand by virtues. I will propose a fairly naturalist view of epistemic virtue in the next chapter.

## Notes

1. A version of this chapter's argument can be found in (Ohlhorst, 2021).
2. In (Ohlhorst, 2022), I argue that this exemption from epistemic evaluation is mistaken.
3. Indeed, coherentism is the internalist notion of justification *par excellence*.
4. I have argued elsewhere (Ohlhorst, 2019) that there are extreme cases of self-deception that are based on hinges.

5. Pedersen (2022) defends exactly this view.
6. Wright himself raises this issue only for entitlement of substance because he does not think that it extends to entitlement of cognitive project or to strategic entitlement. However, given that the issue here is structured like the problem of demarcation and because I do not see a fundamental difference between bizarre ontologies and bizarre methods (which yield bizarre ontologies), I prefer to try to solve the issue wholesale rather than piecemeal.
7. Nikolaj Pedersen (2020) makes a move in this direction, arguing that entitlement becomes obsolete if it points at a value beyond itself; see Chapter 5.
8. Compare this to Martin Kusch's (2019, p. 273) notion of *non-neutrality*. This is a very general property that includes evidential underdetermination.
9. Indeed, Wright argues that sceptical arguments do not need to target our first-order justification for our beliefs. They simply undermine our claim to justification (Wright, 2004, p. 210).
10. Pritchard (2016, p. 80) argues in a similar vein against Wrightian entitlement: Trust in P is accompanied by agnosticism about P that undermines the entitlement.

## References

- Boghossian, P. (2006) 'What Is Relativism?', in Boghossian, P. (ed.) *Truth and Realism*. Oxford: Oxford University Press, pp. 13–37. <https://doi.org/10.1093/acprof:oso/9780199288878.003.0002>.
- Briesen, J. (2012) *Skeptische Paradoxa*. Paderborn: Mentis.
- Brueckner, A. (2005) 'Fallibilism, Underdetermination, and Skepticism', *Philosophy and Phenomenological Research*, LXXI(2), pp. 384–391. <https://doi.org/10.1111/j.1933-1592.2005.tb00454.x>.
- Coliva, A. (2015) *Extended Rationality*. Basingstoke: Palgrave Macmillan.
- Coliva, A. and Moyal-Sharrock, D. (eds.) (2016) *Hinge Epistemology*. Leiden: Brill.
- Friedman, J. (2013) 'Suspended Judgment', *Philosophical Studies*, 162(2), pp. 165–181. <https://doi.org/10.1007/s11098-011-9753-y>.
- Hazlett, A. (2014) 'Entitlement and Mutually Recognized Reasonable Disagreement', *Episteme*, 11(1), pp. 1–25. <https://doi.org/10.1017/epi.2013.46>.
- Kuhn, T. (1996) *The Structure of Scientific Revolutions* (3rd ed.). Chicago: The University of Chicago Press.
- Kusch, M. (2019) 'II – Relativist Stances, Virtues and Vices', *Aristotelian Society Supplementary Volume*, 93(1), pp. 271–291. <https://doi.org/10.1093/arisup/akz007>.
- Lakatos, I. (1976) *Proofs and Refutations*. Cambridge: Cambridge University Press.
- McGlynn, A. (2017) 'Epistemic Entitlement and the Leaching Problem', *Episteme*, 14(1), pp. 89–102. <https://doi.org/10.1017/epi.2015.63>.
- Montmarquet, J. (1992) 'Epistemic Virtue and Doxastic Responsibility', *American Philosophical Quarterly*, 29(4), pp. 331–341.
- Moyal-Sharrock, D. (2004) *Understanding Wittgenstein's on Certainty*. Basingstoke: Palgrave Macmillan.
- Ohlhorst, J. (2019) 'How to Tragically Deceive Yourself', *Les ateliers de l'éthique*, 13(2), pp. 48–69. <https://doi.org/10.7202/1059499ar>.

- Ohlhorst, J. (2021) 'Epistemic Austerity: Limits to Entitlement', *Synthese*, 199(5–6), pp. 13771–13787. <https://doi.org/10.1007/s11229-021-03397-w>.
- Ohlhorst, J. (2022) 'Is There a Problem of Demarcation for Hinges?', *International Journal for the Study of Skepticism*, 12(4), pp. 317–330. <https://doi.org/10.1163/22105700-bja10040>.
- Pedersen, N. J. L. L. (2009) 'Entitlement, Value and Rationality', *Synthese*, 171(3), pp. 443–457. <https://doi.org/10.1007/s11229-008-9330-x>.
- Pedersen, N. J. L. L. (2020) 'Pluralist Consequentialist Anti-Scepticism', in Graham, P. J. and Pedersen, N. J. L. L. (eds.) *Epistemic Entitlement*. Oxford: Oxford University Press, pp. 297–326. <https://doi.org/10.1093/oso/9780198713524.003.0011>.
- Pedersen, N. J. L. L. (2022) 'Entitlement, Generosity, Relativism, and Structure-Internal Goods', *Metaphilosophy*, 53(4), pp. 486–511. <https://doi.org/10.1111/meta.12557>.
- Popper, K. (1962) *Conjectures and Refutations*. New York City: Basic Books.
- Pritchard, D. (2016) *Epistemic Angst*. Princeton: Princeton University Press.
- Reid, T. (2012) *An Inquiry Into the Human Mind*. Cambridge: Cambridge University Press.
- Vogel, J. (2004) 'Skeptical Arguments', *Philosophical Issues*, 14(1), pp. 426–455. <https://doi.org/10.1111/j.1533-6077.2004.00037.x>.
- Wittgenstein, L. (1969) *Über Gewissheit – On Certainty*. Edited by G. E. M. Anscombe and G. H. von Wright. Oxford: Blackwell.
- Wright, C. (2004) 'I – Crispin Wright: Warrant for Nothing (and Foundations for Free)?', *Aristotelian Society Supplementary Volume*, 78(1), pp. 167–212. <https://doi.org/10.1111/j.0309-7013.2004.00121.x>.
- Wright, C. (2014) 'On Epistemic Entitlement (II)', in Dodd, D. and Zardini, E. (eds.) *Scepticism and Perceptual Justification*. Oxford: Oxford University Press, pp. 213–247. <https://doi.org/10.1093/acprof:oso/9780199658343.003.0011>.
- Young, J. O. (2018) 'The Coherence Theory of Truth', in Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2018). Available at: <https://plato.stanford.edu/entries/truth-coherence/#Ver>.

# 4 Virtue

## 4.1 Introduction

In the preceding chapters, I argued that non-evidential believing and warrant are real and epistemologically important phenomena. However, in the last chapter, it also became clear that the account has significant gaps, notably concerning the problem of demarcation. Consequently, we need an account of epistemic value that can bridge the problem of demarcation. That is, it needs to reduce epistemic arbitrariness to avoid issues concerning the arbitrariness of beliefs or a rampant relativism. I suggested that epistemic virtues can do that job. Epistemic virtue shifts the locus of epistemic evaluation from the single belief to the epistemic agent. The problem of demarcation arises from the presupposition dependence of our beliefs; therefore, a shift to the agent may be helpful.<sup>1</sup>

What are virtues? In the most general sense, they are dispositions to behave excellently. There are, for example, the virtues of artefacts: The virtue of a clock is to tell the time reliably under all sorts of circumstances, and a virtue of a painting is to leave a lasting impression on its viewer. There are also the moral virtues of agents: Courage is the virtue of overcoming fear in dangerous circumstances, and generosity is the virtue of giving to others who are in greater need than yourself what you can spare.

We are interested in epistemic virtues here, that is, an agent's dispositions towards behaving *epistemically* excellently. Traditionally, there have been two independent accounts of epistemic virtue and each interprets epistemic excellence differently.

The first, virtue reliabilism, takes excellence to mean reliably producing true beliefs. On this account, virtues are an agent's dispositions towards reliably producing true beliefs. These reliabilist virtues notably include faculties like eyesight or our reasoning competence. A reliable faculty is an agent's excellence. Therefore, the faculty's deliverances, reliably true beliefs, can be *credited* to this agent – for example, most of the time that an agent sees something, she acquires knowledge because she employs a reliable faculty.

Hence, she can be considered as possessing the virtue of sight. When you rely on your faculties, you trust the hinge that they are reliable.

The second, more Aristotelian, account of epistemic virtue interprets excellence as being an excellently motivated habit. This account is called virtue responsibilism. These virtues are indirectly motivated: The dispositions in question are epistemic habits that have been acquired out of a love of truth, which is the excellent motivation.

An example for a responsibilist virtue is *judiciousness*, the habit of not jumping to conclusions. The habit may be acquired with the motivation of not wanting to believe things with an insufficient foundation. Another example is intellectual courage, the virtue of pursuing an investigation out of a love of truth even in the face of external resistance. Intellectual courage also emphasises that responsibilist virtue is closely related to or cannot be properly separated from moral virtues. Namely, intellectual courage might be an epistemic *variant* of the moral virtue of courage. *Vice versa*, moral virtues require intellectual virtues to guide them. For instance, courage requires good judgement about the imminent danger to be dealt with. Responsibilist virtues steer how we are cognitively active and how we pursue cognitive projects, therefore they influence the hinges we need for cognitive projects.

These two accounts were originally thought to compete with each other and to be incompatible. Reliabilists accuse responsibilism of overintellectualising epistemic states, while responsibilists accuse reliabilism of being too reductive. I believe that both criticisms are correct: Neither virtue theory gets human epistemology entirely right. In this chapter, I will argue that instead of competing, the two accounts of virtue describe two complementary *types* of virtue.

Many virtue epistemologists have come around to considering reliabilism and responsibilism as compatible. I call this trend virtue epistemological ecumenism. Heather Battaly (2015) suggests a virtue, epistemological pluralism. John Greco (2010) proposes an agent reliabilism where the different virtues account for different epistemic states. Ernest Sosa (2015) argues that we should consider virtues to be competences of an agent rather than just reliable faculties.<sup>2</sup>

My own proposal will track both Battaly's pluralism and Greco's and Sosa's agent reliabilism but it expands on them. The usual ecumenist approaches to render the virtues compatible are teleological. That is, they make the virtues compatible by drawing on the shared epistemic norms that all virtues contribute to. I argue instead that virtues are unified by their foundations. I have called virtues excellent dispositions and I aim to unify the epistemic virtues through the *foundations of the dispositions* instead of their excellence.

I shall argue that the dispositions grounding our virtues are the dispositions of our cognitive apparatus. Interestingly, this apparatus has several ways of functioning. *Dual process theory*, a prominent account in cognitive psychology, argues that our cognitive functioning works in two

different modes: Type 1 processes are fast, automatic, and context-specific; meanwhile Type 2 processes are slow, controlled, and general (Kahneman, 2011; Evans and Stanovich, 2013).

These two types show remarkable parallels to reliabilist and responsibilist virtues, respectively. I will argue that reliabilist virtues are virtues of Type 1 cognition, while responsibilist virtues are virtues of Type 2 cognition. I shall call them Type 1 virtues and Type 2 virtues, respectively.

In this chapter, I will first introduce the two theories of epistemic virtue and argue why they each independently fail to account for the whole breadth of human cognition. I will argue that the two might fruitfully be brought together on the basis of our cognitive psychology. First, I lay out the parallels between the two kinds of virtue and the two types of cognition. Second, I develop a detailed account of each type of virtue as a cognitive excellence of a type of cognition.

## 4.2 Epistemic virtues

I called virtues dispositions to behave excellently. This means that virtues have two parts: They are *dispositions* which tend to lead to *excellent behaviour*. Fragility is a typical example of a disposition: It is the disposition to shatter under the application of sufficient force. Each disposition has a ‘seat’ (Sosa, 2015, p. 27), namely, the structure that accounts for the disposition. For fragility, it is the fragile object’s molecular structure.

Epistemic virtues are dispositions to process information and form beliefs. Hence, they are much more complex dispositions than fragility. Their seat is not simply laid down in the shape of our skulls. Instead, epistemic virtues are seated in our psychological and biological make-up.

The topic of the greatest disagreement between virtue epistemologists, however, concerns the epistemic excellence involved. This debate about what the excellences are, that our virtues produce, leads to broader epistemological debates about the nature of epistemic value and goodness.

Virtue reliabilists argue that the epistemic good we aim for is the production of true belief. A disposition to reliably produce true beliefs is a virtue. Meanwhile, virtue responsibilists argue that the excellence involved is a love for truth. Here, a disposition guided by a desire for truth is a virtue – a respect for truth in Sylvan’s (2020) terms. This section offers an introduction to the virtue epistemological debate; readers acquainted with the issues involved can skip it.

### 4.2.1 *Virtue reliabilism*

Traditional virtue reliabilism is inspired by process reliabilism (Goldman, 1999), hence the name. Like process reliabilism, virtue reliabilism subscribes to instrumental veritism, the thesis that the only *fundamental* epistemic value relevant to virtues is gaining true beliefs and avoiding false

ones. Other epistemic states are only valuable insofar as they contribute to the production of true beliefs.

A disposition is excellent on a reliabilist view if and only if it produces more epistemic value than disvalue. *Prima facie* that means, if and only if a trait produces more true than false beliefs then it is reliable, therefore excellent, and above a certain threshold of reliability also a reliabilist virtue.<sup>3</sup>

This gives you the basic idea, but there are many limitations to this view. It is obviously a simplification: A disposition may be a virtue by directly and reliably producing more true than false beliefs,<sup>4</sup> but it may also be indirectly virtuous by making the other faculties more reliable. That is, dispositions may also play the instrumental role of allowing or helping to bring about more true beliefs. Such traits are called auxiliary virtues (Sosa, 2015, p. 61). A habit of doing every calculation twice, for example, does not in itself produce true beliefs, but it supports our calculating faculty in producing more true beliefs.

What are the dispositions that reliably produce true beliefs, according to virtue reliabilism? We could propose that *any* disposition that produces more true than false beliefs is a reliabilist virtue. That is, we could individuate virtuous dispositions exclusively through their reliability and refuse to impose any other constraints on what counts as a virtue. This would, however, lead straight to the so-called generality problem because we could gerrymander anything into a reliable faculty; for example, a disposition to recognise yucca palms on Saturdays. The generality problem requires that virtues be individuated in a ‘principled’ manner. Simply individuating virtues by grouping the reliable processes into a disposition would not be principled (Conee and Feldman, 1998). Virtue reliabilists have therefore proposed different accounts of what kinds of dispositions the virtues are.

The original version of virtue reliabilism was also called faculty reliabilism. On this account, the virtues’ dispositions are individuated to *faculties* like vision, hearing, or memory. That is, they are our biologically evolved capacities that deliver beliefs from environmental or other information. If these mechanisms work reliably, as they often do, then they are virtues. Given the archetype of perceptual faculties, faculty reliabilism focuses strongly on innate capacities.

These faculties are dispositions with a simple input–output structure. That is, given some informational or evidential input, the stable faculty will deliver some belief as an output. The output is evaluated according to the veritist norm: If the faculty reliably delivers more true than false beliefs, then it is a reliabilist virtue. When we rely on a reliabilist capacity, then we trust the rule that it actually is reliable.

This simplicity, however, is also virtue reliabilism’s greatest weakness. Many virtue theorists think that reliabilism is much too reductive to account for agents’ epistemic excellences. Reliability does not do justice to the notion of virtue. An epistemically virtuous agent is more than just reliable.



Additionally, Sarah Wright (2010) argues that epistemic virtue does not even *require* reliability. She argues that some cause *c*'s raising the likelihood of a rare event *e* is enough to credit *e*'s occurring to *c*. Meanwhile, *e* may nevertheless be exceedingly rare, thus undercutting reliability. Classifying a new species, for example, is an epistemic achievement of this kind, which may be credited to an agent's virtues but does not require reliability in discovering new species.

The understanding of some phenomenon, let us say the behavioural dynamics on social media, is another manifestation of epistemic virtue. But again, this epistemic achievement seems to have little to do with reliability: There is no faculty that reliably delivers an *understanding of social media*. Such an understanding is too complex to be simply acquired by a series of reliable processes. It is not achieved reliably, but through diligent research and study.

In general, we can say that reliability on its own cannot account for the whole range of excellent human cognition. Hence, reliabilist virtues on their own are not enough. As a consequence, reliabilism has evolved quite a bit and shifted from this focus on reliable faculties to focus more on the so-called *agent reliabilism*. Greco (2010) argues that reliabilist and responsibilist virtue deliver different types of justification, both of which are necessary for full knowledge. Sosa (2015) argues that the dispositions involved in reliabilist virtues are *competences*. Competences as virtues have a different structure than faculties: They are dispositions to succeed at something if you try to do it.

The typical example is archery as a competence. If an archer tries to hit the target and does so successfully then she manifests a competence. In epistemological terms, this can be translated into the claim that if an agent successfully tries to form a true belief, then she manifests a competence and thus epistemic virtue. I will return to some agent reliabilist accounts later in this chapter.

It is because human cognitive achievements are under-described by reliabilist faculty virtues that virtue epistemology has also developed in another direction, namely responsibilist virtue epistemology, which argues that intellectual agency needs to be guided by Aristotelian virtues.

#### 4.2.2 *Virtue responsibilism*

While reliabilist virtues are modelled on faculties like vision, responsibilist virtues as first conceived of by Lorraine Code (1987) and James Montmarquet (1987) take Aristotelian virtues as their model. That is, responsibilism tries to track virtue ethics more closely.

On the Aristotelian view, a virtue is a disposition to manage one's emotions well (Aristotle, 2004). Hence, this involves a two-part diagnosis: One's

emotional dispositions and the normative goal of how to deal with them. The goal on this view is to hit the right mean between opposed emotions. So bravery is the disposition to feel and act in the mean between fear and recklessness. The disposition obviously needs to be stable, that is, it must function across contexts, hitting the right mean in all situations. According to Aristotle, such a disposition is acquired by habituation, that is, by acting bravely in appropriate situations while avoiding recklessness in too dangerous situations and avoiding cowardice in situations that can be overcome.

You might now wonder what the management of our emotions, however laudable it may be, has to do with epistemic excellence. Epistemology is not usually considered as being the domain of emotions, but rather of cognition. But as I mentioned, virtues are not blind, they are context-sensitive. In order to be brave, you need to successfully evaluate how dangerous a situation is and whether the danger can reasonably be overcome as well as what is to be gained from attempting to overcome it. For that purpose, Aristotle introduced intellectual virtues: *Sophia* and *sophrosyne*, which are usually translated as wisdom and practical wisdom.

While he wrote relatively little about these virtues, they are conceived of as dispositions that govern the other dispositions. Responsibilist virtue epistemologists have explored this idea further. The key characteristics of responsibilist virtue epistemology are: A focus on a good motivation, acquired habitual character traits, and the important role that these virtues play for morality.

Let us then return to the normative dimensions of responsibilist virtue epistemology. The key aspect here is the *epistemically good motivation*. Responsibilist virtues are taken to be characterised by, for example, a love of truth (Montmarquet, 1992). That is, if the disposition is borne out of a love of truth, then it is a responsibilist virtue. The idea here is that this epistemically excellent motivation guides the agent in acquiring better epistemic habits: To be more judicious, impartial, conscientious, etc.

A second normative, though not strictly epistemic, aspect is that responsibilist virtues are essentially bound up with moral virtues; something we saw earlier with respect to the guiding function of Aristotelian intellectual virtues (Zagzebski, 1996). A further idea is that intellectual virtues are also already in part moral virtues: If you are fully intellectually virtuous it is impossible or very hard to be morally vicious. This connects the epistemically good motivation, the love of truth, to a morally good motivation.<sup>5</sup> That is, it could be that epistemic virtue not only requires a love of truth but also a will to do right.

Meanwhile, virtue responsibilism is quite vague with regard to dispositions. The dispositions constituting responsibilist virtues are considered to be *habits*, that is, tendencies to do certain things because you did similar things before. This practical description is also very broad. It gives no

further fundamental details on the underlying mechanisms. As soon as repetition produces a behavioural disposition that becomes manifest in certain contexts, it can be considered a habit.

As with reliabilist dispositions, we may individuate anything as a habit if we just gerrymander the context in which the behaviour is produced. You might, for example, postulate a habit of taking walks on sunny spring Sundays in leap years. An agent *may* have such a dispositional profile, but it is arguably not really grounded in a genuine habit and guided by the motivation to take walks on sunny spring Sundays in leap years but is instead arbitrarily gerrymandered.

The important thing for the responsibilist here is that a habit can be acquired and trained. You can do something epistemically virtuous, for example, double-checking on a first impression, for the first time; and then do it again and again. Once you have acted in that way sufficiently often in the appropriate circumstances, it becomes effortless. You have formed a habit and this habit may be part of the virtue of diligence.

Such a habit will be a virtue if it is underwritten by an excellent epistemic motivation, that is, if the process that led to its acquisition as well as the way the agent sustains it are underwritten by an excellent epistemic motivation, a love of truth. When you are virtuous, you do not double-check because you are scared of others' judging that you made a mistake, but because you genuinely cared and still care about getting it right.<sup>6</sup> Your virtuous habit is based on your epistemically excellent motivation. Similarly, you sustain the habit because you still care about truth. If you were to find out that double-checking is actually detrimental to your epistemic performance, you would try to shed your habit or to transform it in some way.

Just as reliabilist virtues fail to account for a part of human cognitive achievement, responsibilism too has its blind spots. Responsibilist virtue is a highly intellectualised type of virtue. It concerns the search for truth, the structure of investigation, and so on – it is about how we pursue cognitive projects. But not all knowledge is so sophisticated. Most of the things we know do not require such a demanding notion of virtue.

Consider the requirement of a desire for truth. We learn a great many things without any motivation, it just happens to us. I cannot help but hear the noise a car makes – even if I am not in the slightest interested in it. What sort of virtuous habit should be at work in the acquisition of the knowledge that 'this car is loud'?

This over-intellectualisation creates a further worry: Only few actually possess such demanding intellectual virtues. Take, for example, children – they likely lack a notion of truth sufficiently developed to desire it, and even if they do possess such a notion, they probably would not have come to form stably virtuous dispositions directed at truth. If good epistemic standing were to exclusively be explained through responsibilist virtues,

it would be exceedingly rare. But children in particular can achieve very impressive cognitive feats without even being aware of it. Consider, for example, how fast and well they learn languages. Hence, the responsibilist virtue approach has gaps too, notably as concerns automatic non-agential cognitive achievements.

### 4.3 Unification

Consequently, both reliabilism and responsibilism fail to give us a complete account of cognitive excellence or virtue. This is known as the scope problem (Pedersen, 2017, pp. 52–53). The scope problem in general is that there is a certain range of phenomena that we would like to call X, but none of the available accounts of X can accommodate all of the phenomena. In epistemology, the notion of justification has been subject to such a treatment in Alston (1985), Goldman (1988), and Burge (2003).<sup>7</sup> The scope problem with respect to virtue is that it should account for both quotidian epistemic achievements and more extraordinary epistemic feats. Reliabilism and responsibilism, respectively, can only fully accommodate and provide suitable explanations of one of these aspects.

My solution is to unify the two. More exactly, I will show how responsibilist and reliabilist virtue are complementary because they describe the cognitive excellence of different *types of information processing*. Reliabilist faculties cannot account for distinctly human excellent cognitive achievements, for example, new discoveries. Meanwhile, responsibilist habits are much too demanding and over-intellectualise some human cognitive achievements; innate perfect pitch, for example, has nothing to do with good motivation.

However, if we adopt both traditional accounts at the same time then we can cover the whole range of human cognition. This raises the question of how the two fit together. In the last decade, the project of unifying or rendering compatible responsibilist and reliabilist virtue approaches has gained considerable traction. Most of these ecumenist projects focus on unifying virtues by focusing on their normative goals. That is, they try to find a single higher norm that both responsibilist and reliabilist virtues adhere to in their own way. I call this teleological unification.

Unifying reliabilism and responsibilism through a shared norm leads to an unsatisfying reductionism: One of the virtue type's epistemic goodness is reduced to virtues of the other type's goodness. I will instead try another avenue: I focus on the *dispositional* aspect of what a virtue is and give an account of how the dispositional parts of reliabilist and responsibilist virtues are unified. You may have noticed that the reliabilist and responsibilist accounts are fairly vague regarding the nature of the dispositions. According to reliabilists, these are faculties with an input–output structure<sup>8</sup>; according to responsibilists, these are habits.

I want to focus on the substrate of these dispositions. When we talk about virtues, we actually talk about human virtues. This means we can examine the sort of human capacities in which the virtues are anchored. I want to account for epistemic or intellectual virtues. I think these are anchored in our cognitive system, which is responsible for our epistemic achievements. This means that I have to look to cognitive psychology to explain what virtues are. As it happens, one of the most popular<sup>9</sup> theories in the field is dual process theory, which has some striking parallels to virtue epistemology.

#### 4.3.1 *Dual process theories*

Dual process theories argue that there are two types of cognitive processes.<sup>10</sup> These two types more or less track the folk distinction between intuition and explicit reasoning (Kahneman and Frederick, 2002, p. 51). That is, the distinction we make between beliefs we just find ourselves forming and beliefs we acquire through a process of explicit reasoning.<sup>11</sup>

Type 1 cognition corresponds to the folk notion of intuition. Its hallmark is that it occurs without any guidance: Information is automatically processed and its output delivered. It is very efficient, fast, and requires no effort. These efficient Type 1 processes fulfil a very specific function or algorithm and always only execute that function. The task that a Type 1 process thereby solves may nevertheless be highly complex. Face recognition is a paradigmatic Type 1 process: It can successfully differentiate and recognise individuals on the basis of minor cues, (Osman, 2004) and it develops early in infancy (Carey, 2009, p. 176). However, if there are no faces present, the responsible process type would remain at best inactive, at worst it would find faces where there are none.

Arguably there is nothing that unifies Type 1 cognition neurologically; it rather consists of a set of disparate processes that all exhibit the aforementioned traits. What allows them to be grouped into a single type is how they differ from the more unified Type 2 processes (Evans, 2008, p. 270).

Type 2 cognition would be called ‘explicit reasoning’ in folk terms. ‘Explicit’ here by no means implies linguistic reasoning, it can also be carried out through visual imagination, for example. It is rather characterised by the degree of control we have over it. We can engage it at will and steer its direction. However, it demands concentration and is very slow. When compared to Type 1, Type 2 is very inefficient.

Type 2’s inefficiency does not, however, mean that it is useless. Given that we can engage in it at our leisure and that it does not have a fixed input–output structure, we can start with any information. That implies that it can be used to try to solve any problem and that it functions independently of context. Its relative inefficiency is the price of its universal applicability.

To stay with faces: A renaissance painter may reason about the angles and ratios of a face, trying to find a face’s ideal proportions by engaging

her Type 2 reasoning. However, she could use these same processes equally well to represent the intricacies of a neoclassicist façade.

Table 4.1 gives an overview of the key characteristics of Type 1 and 2 cognition. My favoured theory of what underlies this difference is that Type 2 cognition involves the use of working memory, while Type 1 cognition lacks this element (Samuels, 2012, pp. 141–143; Evans and Stanovich, 2013, p. 225). The idea is that in Type 2 cognition information is not processed *en bloc* but split up in parts and manipulated sequentially, while the rest is stored in short-term memory. For example, Type 1 processes a face as a whole, while Type 2 will operate on its parts one after the other.<sup>12</sup>

Distinguishing these process types serves a wide range of purposes in cognitive psychology. One of the classic instances of the phenomenon is in the psychology of learning: There is an explicit and demanding Type 2 learning process, the sort of thing you engage when you study vocabulary in a new language. However, there are also implicit learning processes running in parallel to this: Subjects learn to (unconsciously) recognise and respond to patterns. This is assigned to Type 1 cognition (Frankish, 2010, p. 919).

The other great classic application of dual process theory is reasoning. Subjects may know logical or probabilistic rules, understand them, and know how to apply them in the abstract. But when confronted with concrete instances of, for instance, *modus tollens*, they tend to systematically fail, as in the Wason selection task. Similarly, people will commit the conjunction fallacy given that their Type 1 cognition outputs the belief that ‘fitting’ conjunctions are more probable than a non-fitting single conjunct of that ascription. Given a description of Linda as participating in protests and an activist, most people will think it is more likely that she is a feminist bank teller rather than *just* a bank teller. But by entailment the latter is at least as likely as the former. The theory is that subjects use a Type 1 heuristic at the same time as they go through explicit Type 2 reasoning. Depending on the circumstances, the output of the one trumps the other (Kahneman and Frederick, 2002).

Type 1 and 2 processes describe different ways in which humans cognise. The two types interact continuously by providing each other with

Table 4.1 System 1 and System 2 cognition (*cf.* Stanovich and West, 2000, p. 659)

<i>System 1</i>	<i>System 2</i>
Associative	Rule-based
Holistic	Analytic
Automatic	Controlled
Relatively undemanding of cognitive capacity	Demanding of cognitive capacity
Relatively fast	Relatively slow
Acquisition by biology, exposure, and personal experience	Acquisition by cultural and formal tuition

information, competing with each other, and controlling each other. Almost no complex human cognitive endeavour is purely of one type; that is, they complement each other and together make up human cognition.

#### 4.3.2 *Parallels*

Dual process theories have played a role in epistemology, although it is smaller than one might expect. Most notably, Keith Frankish (2004) has developed a dual process theory of belief. But mostly, dual process theories are used in the epistemology of reasoning and rationality.<sup>13</sup> I suggest a different approach: The virtue types track the two process types.

If you consider Type 1 cognitive processes and reliabilist virtues, there are striking parallels. Both are conceived of as being able to function independently of voluntary guidance, both are very context-specific, and both have a focus on efficiency.

Type 2 processes and responsibilist virtues also have striking parallels. Both involve an element of voluntary control, both manifest a sensitivity to reasons and, finally, both may be acquired and trained.

I believe that these parallels warrant examining how a dual process account of reliabilist and responsibilist virtues would play out. This holds especially because it would give us a novel way of rendering the two approaches compatible: Reliabilist virtue would relate to responsibilist virtue while Type 1 cognition would relate to Type 2 cognition. The thesis I will examine in the next sections is that epistemic virtues are excellences of our cognitive apparatus. Given that our cognitive apparatus operates with two types, we would also expect two types of virtues, reliabilist Type 1 virtues and responsibilist Type 2 virtues.

There are different ways of modelling the relationship between virtue epistemology and dual process theory. In ascending order of the strength of the relationship, the first is that reliabilist virtue is *analogous* to Type 1 cognition, and responsibilist virtue is analogous to Type 2 cognition. This would point to the role of intuition and reason, respectively, in the analogues. The second account essentially consists in what I argued in this section. Namely, that Type 1 cognition has significant correlations with the exercise of reliabilist virtues and Type 2 cognition correlates with the exercise of responsibilist virtues, but there might also be exceptions. I want to advocate for the third option, which is the strongest, namely, that reliabilist virtue is essentially the virtue of Type 1 cognition while responsibilist virtue is essentially the virtue of Type 2 cognition. This naturalises the two virtue types by clearly indicating their roots in parts of the human organism's cognitive apparatus.

This, then, would provide an account of the dispositional aspect of the virtues. Given that the two cognitive types have different functions, a

further interesting consequence is that their virtues may have different normative profiles. We can explain the different aspects of what counts as an excellence of each type from out of these diverging functions. This is what Allan Hazlett (2016, p. 264) would call an *energetic account of virtue*.

## 4.4 Dual process virtues

### 4.4.1 Reliabilist Type 1 virtues

I argued that virtues in general are dispositions towards excellence. For my claim that reliabilist virtues are virtues of Type 1 processes, this means that they are dispositions of Type 1 processes towards epistemic excellence. Hence, I have to explain what Type 1 dispositions are and how they can be epistemically excellent.

Type 1 cognition is thought to be made up of a manifold of specialised cognitive capacities. These operate by activating at a certain kind of input; for example, auditory signals with certain characteristics. These Type 1 modules then process the input according to a fixed pattern and deliver some informational output; for example, the recognition of some sounds as a major chord. Another important example of our Type 1 modules are the dedicated input analysers of our core cognition (Carey, 2009).

I take these capacities to be the dispositions in question. More abstractly, they are dispositions to deliver beliefs in certain initial conditions. This is a very fine-grained account of reliabilism. Often, the faculties or dispositions of virtue reliabilism are carved up more coarsely, for example, as the auditory faculty in general.

We have many Type 1 capacities. Each potentially constitutes one reliabilist virtue. They are stable structures and very efficient. These capacities run automatically and more or less unconsciously. We are only aware of their output while the computation that went into delivering these results remains hidden from us.

These, then, are the dispositions we are examining. What makes a given capacity epistemically excellent and thus into a virtue? Given their fairly simple structure, I believe that epistemic excellence for Type 1 capacities can only mean reliability. That is, the only epistemic norm to which Type 1 processes are subject is that they deliver as many true beliefs as possible and avoid as many false beliefs as possible. Essentially, these processes transform some informational input into a belief. What other epistemic values could we desire from Type 1 processes apart from their being accurate and avoiding error? When we rely on a Type 1 capacity, this is also the rule that we trust in its regard.

Type 1 processes run automatically and unconsciously. While they may be sensitive to evidence in a broad sense, their output belief does not come



with the reasons behind its formation. Type 1 thus cannot satisfy higher rational requirements like reflective access.

A demanding modal profile is also unavailable to such processes as they function in specific contexts. Type 1 processes are often a product of evolution and they are thus adapted to function (more or less reliably) in the organism's normal environment, that is, environments similar in a relevant sense to the environment in which the organism evolved. In non-normal environments, Type 1 cognition is out of its depth and quickly becomes unreliable. Note that we as humans have transformed our own environments to such a degree that they are often abnormal – just consider that you will most likely be reading this text from an array of glowing dots. Human visual processes did not evolve to look at screens – thus computer screens have to be designed to mimic our normal environment – but this only goes so far and looking at screens for prolonged periods is exhausting. Type 2 cognition, however, is modally more robust and not bound to specific normal environments. The excellence that Type 1 cognition may deliver is reliability in environments sufficiently similar to the organism's evolutionarily normal habitat.

There are two aspects to the reliability of a Type 1 capacity. The first is the range of contexts across which the capacity can function reliably – call this the capacity's strength. The second aspect are the contexts in which the capacity actually runs and delivers beliefs – this is called the capacity's trigger reliability.

Type 1 processes are highly specialised. Most Type 1 processes only function reliably in a narrow range of contexts. Beyond those contexts, they become unreliable. This means that we can compare the number of contexts in which different capacities function reliably. The more contexts in which a capacity reliably delivers true beliefs, the *stronger* it is. Strength describes the range of contexts in which the process is reliable.

The other key issue regarding the virtue status of these Type 1 capacities is *when* they fire. Given their context-specificity, capacities are only reliable in very specific contexts – even a strong capacity has its limits. Type 1 cognition is not domain-general. Therefore, a key aspect of virtuousness in a Type 1 capacity is that it activates and processes information only in those contexts where it is effective and reliable. Alfano (2013, pp. 151–152) calls this feature a capacity's 'trigger reliability.' If our face recognition were regularly triggered and delivered beliefs about faces in utter darkness, it would be a defective capacity and not a virtue.

There are then two aspects to the reliability of Type 1 capacities: The trigger reliability of their activation conditions and the strength of the process they execute. The single capacity is the potential virtue. This account has great similarities to Jack Lyons's (2019) algorithms and parameters (A&P) view. The algorithm is the process or capacity, while the parameter is the context in which the process runs.<sup>14</sup>

A key point here is that, because the algorithm itself is rigid and inflexible, a Type 1 capacity is reliable only if it fires under appropriate circumstances, that is, circumstances in which the process may actually be successful. My favourite example for this is our face recognition process, which is highly specialised: Sometimes there is ambiguous visual information that may or may not be a face. If that visual input triggers the process, then the Type 1 capacity will form the belief that there is a face even if there was none or *vice versa*. If our face recognition is not kept in check by our other cognitive processes, this may lead to your looking at a photograph without any people in it and forming the belief that there is a ghost in the picture because our face recognition module is activated by some background pattern and falsely recognises it as a face.<sup>15</sup>

There are generally more reliable results if the face recognition process only fires when it is actually reliable. Also note that a face recognition module that has high trigger reliability may turn out to be more reliable than someone else's module that, despite being technically stronger, activates indiscriminately. Discriminating firing means high trigger reliability. For example, a person with a strong face recognition module that can reliably discover faces in many contexts, for example, including very small or blurred faces, but whose face recognition module fires unreliably, may be prone to see faces where there aren't any or not recognise visible faces, because their face recognition activates in the wrong circumstances.

In sum, reliabilist virtue of Type 1 cognition has two aspects: (a) its strength or calculating power, which allows it to more or less reliably extract beliefs from input information and (b) its trigger reliability or activation conditions, which may or may not track the capacity's strength. If (a) and (b) jointly lead to reliable truth delivery and avoid false beliefs, then a capacity is a reliabilist Type 1 virtue. When you rely on a Type 1 capacity, then you trust the rule that outputs of this type are indeed reliable. It is a hinge for you that your Type 1 cognition is trigger reliable and strong under most circumstances.

This means that we have a naturalist account of reliabilist virtue as excellences in our Type 1 processes. These virtues are not the usual suspects; rather they are whatever processes are identified by cognitive psychology as constituting a distinct module or capacity (algorithm).

#### 4.4.2 *Responsibilist Type 2 virtues*

We can develop an analogous notion of responsibilist virtue deriving from the thesis that it designates the excellent functioning of our Type 2 cognition. Type 2 is usually considered as being more unified than Type 1. There are different theses as to what exactly defines Type 2 cognition: Its being rule-based, its being cyclical or recursive, or its involving a working

memory. I think Samuels (2012), and Evans and Stanovich (2013) are right that the latter thesis is the most plausible. That is, Type 2 cognition is the capacity that stores and recalls information in short-term memory. For example, when reading, you need to be able to recall what you just read in order to make sense of the current passage. It also strikes me as plausible to argue that there may be a close relation between our holding information in our working memory and our being aware of it (Samuels, 2012, pp. 141–143; Evans and Stanovich, 2013, p. 225).

Whatever the exact fundamental nature of Type 2 cognition, its characteristics track responsibilist virtue quite well. Let us first consider the dispositional side: Type 2 cognition does not consist of input–output mechanisms, it is more open-ended and agent-guided. Consider your trying to find a logical proof for some theorem: While you may follow a strict algorithm that inexorably leads to the proof,<sup>16</sup> you may also instead go about it more creatively and heuristically, for example, by patiently and diligently trying different promising manipulations. The agent is epistemically responsible when using her Type 2 cognition.

Hence, while such a heuristic procedure may be reliable (or not), we can also realise further epistemic values with Type 2 cognition. When you find a logical proof, you do not only reliably acquire truth, you acquire demonstrated certainty and understanding. You can also discover an *elegant* proof, which makes the demonstrated truth much easier to grasp intuitively. Note that you will be more likely to find an elegant proof through heuristic poking than through a strict algorithm. Consequently, Type 2 cognition realises more epistemic values than just true belief.

Returning to the musical example of Type 1 virtues, an epistemic Type 2 activity would be to reconstruct a lost piece of 16th-century music on the basis of some fragments. In order to do this, you need to understand a great many things, hold your fragmentary data in working memory, and be creative but also respectful of 16th-century musical conventions. If you succeed, you will produce a plausible and hopefully (partially) accurate rendition of how the music sounded 500 years ago. Yet accuracy is not necessary in this case, plausibility is sufficient. Thus, Type 2 cognition may give us more epistemic qualities than mere reliability and reliability is not even required for epistemic success in Type 2 cognition, as this old music example shows.

This richer normative profile of Type 2 cognition also translates nicely into what is supposed to be the epistemic excellence of its (responsibilist) virtues. Responsibilist virtues aim at values like understanding (Kvanvig, 2003), internalist justification (Greco, 2010, p. 167),<sup>17</sup> rationality, insight, and moral adequacy (Montmarquet, 1992; Wright, 2010). As argued in Chapter 2, these values may be reducible to a respect for truth (Sylvan, 2020), but not to simple true belief. According to Battaly (2015),

responsibilist virtue is agential, thus it has a more deontological normative profile: If we are to be virtuous, we are required to test our beliefs for defeat (Montmarquet, 1992) and to adopt them only on the basis of sufficient evidence (Clifford, 1999).

These more agential epistemic requirements cannot be satisfied by Type 1 cognition. Naturally, some Type 1 capacities may be functionally equivalent to such features, for example, our visual perception is constituted through many correcting subprocesses. This is probably why we treat visual perception as a kind of folk epistemological gold standard: ‘I don’t believe it until I’ve seen it.’ But visual perception of a nearby well-lit scene is more the exception than the rule for Type 1 processes. Arguably, the fact that some Type 1 processes can track such qualities is the reason why reliabilist and responsibilist achievements are so hard to pry apart: They sometimes look so similar that it seems more like a matter of degree than of kind. The difference in kind, however, arises from the underlying dispositions, not the corresponding normative rules.

This argument shows that by subscribing to Type 2 virtues we do not only subscribe to a pluralism about virtues (reliabilism and responsibilism) but also to a plurality about epistemic values which may or may not be reducible to a respect for truth (Zagzebski, 1996; Wright, 2010; Battaly, 2015). However, while I argue that Type 2 cognition realises additional epistemic values like understanding and discovery which transcend simple true belief (Pedersen, 2017), I do not believe that this normative profile is what classes these virtues among the responsibilist virtues. The underlying Type 2 cognition instead permits realising these supplementary epistemic values.

Type 2 cognition can realise these epistemic values because it is subject to epistemic agency. That is, in a Type 2 process, we can control what we pay attention to, we can look for defeaters for a proposition, and we need to concentrate on what we do. Our Type 2 capacities steer which cognitive projects we pursue and how we go about them. Additionally, Type 2 dispositions or habits are much easier to acquire than Type 1 dispositions<sup>18</sup>: You need to make a habit of paying attention to certain things, asking certain questions, etc.

There is a psychological research programme on such habits called cognitive styles (Kozhevnikov, 2007; Stanovich, 2009), which remarkably also involve traits like diligence, open-mindedness, creativity, etc. In other words, cognitive styles look a lot like the psychological counterpart to responsibilist virtues.

A further nice example of Type 2 dispositions is the logic training that philosophy students go through. It teaches you to isolate single claims and to draw out how they are logically related. If you train well enough, seeing the logical structure of claims becomes a fixed habit. This clearly leads to many epistemic goods: You will understand others’ claims and

arguments better, you will avoid certain kinds of mistakes, you will be able to extract new beliefs from what you already believe and be able to support your own beliefs with arguments.<sup>19</sup> Arguably, when you learn formal logic, you acquire a set of Type 2 dispositions and thus a set of responsibilist virtues.

Consequently, Type 2 virtues are dispositions of our Type 2 cognition to function excellently. These dispositions are under agential control and therefore like habits. They further require an epistemically excellent motivation because otherwise our Type 2 cognition would get distracted by other non-epistemic goals. Type 2 virtues can manifest themselves in pure Type 2 processes, that is, reasoning, or they manifest themselves by guiding and correcting Type 1 cognitive processes. In doing this, they produce a wide range of epistemic values from true beliefs to discoveries, justification, and understanding.

#### 4.4.3 *Reliabilist Type 2?*

In the introduction, I limited my view to faculty reliabilism. But I mentioned a neighbouring view called agent reliabilism (Greco, 1999; Sosa, 2015) which is broader and includes an agent's skills and competences. Such an agent reliabilism could also include our Type 2 capacities and evaluate them with respect to their reliability. An example, that I mentioned earlier, would fall into this category: An agent may have a disposition to reliably prove logical theorems. As mentioned, I also allow that responsibilist Type 2 virtues may be reliable.

The question, then, is that of precedence: Is any reliable disposition also a reliabilist virtue or is this latter determined by the underlying type of cognitive processing? Do we principally focus on the normative side of the definition of virtue or on the dispositional aspect?

To classify all reliable dispositions as reliabilist virtues without saying anything about the nature of their underlying dispositions would be to open the gates to the generality problem. The issue here is that if we individuate virtues only normatively, that is, by their reliability, they will not be individuated in a principled way (Conee and Feldman, 1998, pp. 3–4). Without any restrictions on what may count as a disposition, we can gerymander anything into a reliable disposition.

In the most extreme of cases, we could postulate a virtue to believe truly, which would be the arbitrary disposition consisting of all particular events that *de facto* lead to a true belief. Individuating virtues teleologically, that is, by grouping true-belief-producing events into a virtuous dispositional type, puts the explanatory cart before the horse – the only thing we learn from such an arbitrary account is that we value reliability. Individual belief-forming processes could be assigned to any arbitrary type.

This means we need to draw some clear lines concerning what counts as a genuine disposition. On my count, what makes processes into a type is that they belong to the same cognitive capacity. This process type is then evaluated for its strength and trigger reliability. If we want to examine whether some capacity is reliable, we examine *all* its instances.

The nice thing about Type 1 dispositions is that they are well-individuated and modularised: There is one algorithm or capacity, and reliability arises from whether the process actually runs when it is effectively reliable. With Type 2 cognition, there is no such simple process type that we can point to. Individuation is much more difficult: What is the reliable Type 2 process or disposition that runs when we try to prove some theorem? Type 2's dispositions do not fit neatly with Type 1 reliabilist virtues. We do not actually individuate a genuine type by simply grouping reliable Type 2 processes. Additionally, given that Type 2 processes are under agential control like responsibilist virtues and the fact that responsibilist virtue may include reliability (Zagzebski, 1996), I believe that on balance it is more elegant to assign reliabilist virtues to Type 1 cognition and responsibilist virtues to Type 2.

I do not aim to convince the radical virtue reliabilist who takes responsibilist virtue to be unimportant or uninteresting. Instead, I aim to propose a joint-carving view of reliabilist and responsibilist virtue that can appeal to philosophers who think both types of virtue are important.

There are more refined accounts of reliabilism, called agent reliabilism, that argue that, although responsibilist virtue plays a role, this is only auxiliary to reliabilist virtues (Lepock, 2011; Fleisher, 2017). These accounts argue that all reliable dispositions are reliabilist virtues, while dispositions instantiating other epistemic qualities will be auxiliary responsibilist virtues in the sense that they support the reliability of the primary reliabilist virtues. I think that this auxiliary function is only part of the normative profile of a Type 2 virtue.

On this account, there would be reliable and unreliable Type 2 dispositions. If the latter contribute to the reliability of the reliabilist capacities, then they make up the responsibilist virtues. Hence, we would have reliabilist Type 2 capacities.

However, this does not mitigate the problem that there are not as distinctly individuated process types or disposition in Type 2 cognition as there are in Type 1 cognition. The latter consists of distinct fairly hard-wired modules, which are characterised by their insulation from other processes. The former appears to be mostly characterised by its use of working memory and sequential processing – a structure that does not lend itself to easily individuating particular processes. But do we really want reliable Type 2 cognitive processes to be classified with modularised reliable Type 1 processes, while unreliable but otherwise epistemically valuable Type 2 dispositions are placed in a distinct category, namely that of responsibilist

virtues, even though they are far more similar to reliable Type 2 dispositions? By this argument, I also defend the autonomy of responsibilist virtues as Type 2 virtues, which are not merely auxiliary to improving our reliabilist virtues' outputs but are sometimes also reliable themselves (*cf.* Baehr, 2011, p. 12).

There is a further agent reliabilist view that will play a key role in Chapter 5: Ernest Sosa's (2015) competence reliabilism. For Sosa, the relevant disposition is a competence, something you are able to do. The norm is whether you succeed when you try to do it; hence, reliability is not specifically geared towards the formation of true beliefs.

Sosa's notion of competence is extremely general. Anything that an agent can do – whether practically or epistemically – is a competence. You can try to see, hear, remember, calculate, walk, reach, throw a ball, find out, demonstrate, plan a trip, go on a trip, and much more.

Sosa's account of competence is highly developed. It gives a clear meta-physical framework concerning what the possessing of a competence consists in. There are three levels: Seat, shape, and situation (SSS) (Sosa, 2015, p. 27). 'Seat' designates the innermost possibility, the structure necessary for the competence in the agent – for example, the seat of our capacity to hear is the anatomy of our ears and so on. 'Shape' designates the variable current state of the agent and whether the agent could currently manifest the competence – for example, being awake is a necessary condition for actually hearing. The 'situation' is the agent's relation to her environment, which is necessary for the manifestation of the capacity – for example, there not being too much background noise.

This accounts for reliabilist virtues; but what role do responsibilist virtues play in this framework? As in Lepock (2011), these have a research-guiding function. Namely, responsibilist competences cannot produce knowledge because they do not reliably produce true beliefs. Instead, they put us *in a position to know*, that is, responsibilist competences make information accessible to our reliabilist competences that then produce the true belief (Sosa, 2015, p. 41).

That means, according to Sosa, that responsibilist virtues are causally necessary for certain kinds of demanding knowledge. But any kind of knowledge also requires a reliabilist faculty. Thus, in the end, Sosa also requires Type 2 reliabilist capacities, because it is very unlikely that Type 2 cognition is incapable of also generating true beliefs.

Sosa's framework relates to my account in the sense that mine is a more developed account of our *cognitive* competences. I give an account of what cognitive competences are and of the cognitive capacities in which they are anchored. What Sosa calls the 'seat' are my Type 1 and Type 2 capacities, Lyons's (2019) algorithms. However, my account does not distinguish 'shape,' which is the agent's inner state, and 'situation' which is the

agent-external context. On my view, this is simply the context, or as Lyons puts it, the parameters, of the core capacity. A virtuous capacity needs to cope with as broad a range of contexts as possible either by being strong and functioning reliably in a broad range of contexts or by being sensitive to its own unreliability, which means possessing high trigger reliability. In my account, it does not make a difference whether the variations in context concern the agent-internal ‘shape’ or the agent-external ‘situation.’

Given that Sosa’s competence reliabilism implies that some reliable Type 2 capacities are reliabilist virtues, I will present a last argument in favour of an exclusively responsibilist Type 2. Clearly, I cannot deny that there are reliable Type 2 competences. However, I claim that the salient explanation for these competences’ reliability lies within the agent’s responsibilist character.

Type 2 cognition lies under the control of the agent and for it to be stably reliable, that is, to actually produce reliably true beliefs, the agent must exhibit responsibilist virtue: Be diligent, inquisitive, etc. That is, on my view, responsibilist virtue does not only create the situation and shape for some reliabilist faculty to manifest itself. Rather, there are responsibilist virtues that by manifesting themselves as responsibilist are reliable in their ‘seat.’ This is possible because responsibilist competences do not only steer other faculties but also themselves. This means that in Type 2 cognition, reliability supervenes on responsibilist dispositions but not *vice versa*, while there are other epistemic values in Type 2 cognition that do not require the production of reliability. Consequently, we will always need to explain excellent Type 2 cognition through its responsibilist features, while reliability is only an additional excellence of some of these responsibilist virtues.

A reliabilist could still think that these reliable Type 2 virtues are reliabilist or she could even deny that such competences exist. The first option softens the notion of reliabilist virtue, although it has a strong framework behind it on Sosa’s account. However, for the most part, it somewhat inelastically groups together virtues that require responsibilist excellences with very basic reliabilist faculties.

The second option appears to be implied in Sosa’s account. This means that responsibilist virtues are pure steering faculties that only govern other reliabilist processes. I think that some accounts of responsibilist virtue may function like that (Baehr, 2011, p. 11), but many claim a more robust role for responsibilist virtues.

My account also falls in the latter camp, given both the focus on agency and on Type 2 faculties. Type 2 cognition has this steering faculty, but it can also steer its own processes: This is what we call ‘reasoning’ in folk terminology. I do not think that reasoning is a reliabilist virtue, rather it can make responsibilist virtue manifest because it can only be reliable if it exhibits responsibilist virtue. Hence, its reliability is constituted by its being responsibilist.



## Notes

1. This chapter is based on Ohlhorst (2022), which contextualises it more in the virtue epistemological debate and also addresses issues like the situationist challenge and the generality problem.
2. Further more or less ecumenical approaches are Zagzebski (1996), Lepock (2011), Axtell (2017), Fleisher (2017), and Mi and Ryan (2020).
3. More recently, Miracchi (2015) and Kelp (2018) have argued that virtue instead aims at knowledge, developing a knowledge-first virtue epistemology.
4. The picture is even more complicated, given that not all true and false beliefs are created equal. Arguably, you also need to weigh beliefs according to their epistemic significance, for example, how many consequences they have (Lepock, 2011).
5. This follows the Platonic identification of truth with the good (Plato, 2005, pp. 88b–89a).
6. Although MacIntyre argues that we may acquire virtues by shifting our motivations in the course of forming the habit (MacIntyre, 1981, pp. 188–189).
7. Greco's (2010) agent reliabilism draws on Goldman's and Alston's notions of justification in his argument about the different types of virtue.
8. This is somewhat misleading, given that the dominant contemporary theory of reliabilist virtue (Sosa, 2015) has a metaphysically rich account of what constitutes a faculty or a competence that grounds a virtue.
9. Dual process theory is not uncontested, but I cannot relitigate all of cognitive psychology. My theory simply assumes that these two cognitive types can be distinguished.
10. Historically, dual process theories did not distinguish between two cognitive types, but systems. However, theorists have come away from the idea that these are separate systems. For my purposes, process types as natural kinds are sufficient (Evans, 2008, p. 270).
11. For introductions to this theme, see, for example, Frankish (2010) and Kahneman (2011).
12. For Type 2 cognition to be useful at all, it therefore needs to be able to operate with variables and placeholders.
13. There are also sketches of dual process accounts of epistemic virtue, for example, by Samuelson and Church (2015), Axtell (2017), and Mi and Ryan (2020) – I examine these views in Ohlhorst (2022).
14. For more detail on the A&P view and a comparison, see Ohlhorst (2022).
15. Obviously, some further bridge assumptions need to be made to get to belief in a ghost; but taking one's face recognition to be reliable is the belief's starting point.
16. In a billion years, because you are not a supercomputer.
17. If Type 2 virtue were limited to delivering internalist warrant, then my account would be a naturalist version of Greco's agent reliabilism.
18. You can anchor a Type 1 disposition in your behavioural profile, but it requires much stricter conditioning. Martial arts, for example, conditions you to acquire certain reflexes that are Type 1, but acquiring such dispositions is different from acquiring a thinking disposition.
19. However, learning logic will not protect you from rationalising your mistaken beliefs. That is, the virtues that formal logic alone teaches will not be enough, but they are an important part of epistemic excellence.

## References

- Alfano, M. (2013) *Character as Moral Fiction*. Cambridge: Cambridge University Press.
- Alston, W. P. (1985) 'Concepts of Epistemic Justification', *The Monist*, 68(1), pp. 57–89.
- Aristotle. (2004) *Nicomachean Ethics*. Edited by R. Crisp. Cambridge: Cambridge University Press.
- Axtell, G. (2017) 'Thinking Twice About Virtue and Vice: Philosophical Situationism and the Vicious Minds Hypothesis', *Logos and Episteme*, 8(1), pp. 7–39. <https://doi.org/10.5840/logos-episteme2017811>.
- Baehr, J. (2011) *The Inquiring Mind*. Oxford: Oxford University Press.
- Battaly, H. (2015) 'A Pluralist Theory of Virtue', in Alfano, M. (ed.) *Current Controversies in Virtue Theory*. Abingdon: Routledge, pp. 7–21.
- Burge, T. (2003) 'Perceptual Entitlement', *Philosophy and Phenomenological Research*, 67(3), pp. 503–548. <https://doi.org/10.1111/j.1933-1592.2003.tb00307.x>.
- Carey, S. (2009) *The Origin of Concepts*. Oxford: Oxford University Press.
- Clifford, W. K. (1999) 'The Ethics of Belief', in Madigan, T. (ed.) *The Ethics of Belief and Other Essays*. Amherst: Prometheus, pp. 70–96.
- Code, L. (1987) *Epistemic Responsibility*. Hanover: University Press of New England.
- Conee, E. and Feldman, R. (1998) 'The Generality Problem for Reliabilism', *Philosophical Studies*, 89(1), pp. 1–29. <https://doi.org/10.1023/A:1004243308503>.
- Evans, J. S. B. T. (2008) 'Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition', *Annual Review of Psychology*, 59(1), pp. 255–278. <https://doi.org/10.1146/annurev.psych.59.103006.093629>.
- Evans, J. S. B. T. and Stanovich, K. E. (2013) 'Dual-Process Theories of Higher Cognition: Advancing the Debate', *Perspectives on Psychological Science*, 8(3), pp. 223–241. <https://doi.org/10.1177/1745691612460685>.
- Fleisher, W. (2017) 'Virtuous Distinctions: New Distinctions for Reliabilism and Responsibilism', *Synthese*, 194(8), pp. 2973–3003. <https://doi.org/10.1007/s11229-016-1084-2>.
- Frankish, K. (2004) *Mind and Supermind*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/cbo9780511487507.004>.
- Frankish, K. (2010) 'Dual-Process and Dual-System Theories of Reasoning', *Philosophy Compass*, 5(10), pp. 914–926. <https://doi.org/10.1111/j.1747-9991.2010.00330.x>.
- Goldman, A. I. (1988) 'Strong and Weak Justification', *Philosophical Perspectives*, 2, pp. 51–69.
- Goldman, A. I. (1999) *Knowledge in a Social World*. Oxford: Oxford University Press.
- Greco, J. (1999) 'Agent Reliabilism', *Philosophical Perspectives*, 13, pp. 273–296. <https://doi.org/10.1111/0029-4624.33.s13.13>.
- Greco, J. (2010) *Achieving Knowledge*. Cambridge: Cambridge University Press.
- Hazlett, A. (2016) 'Epistemic Goods', in Fletcher, G. (ed.) *The Routledge Handbook of Philosophy of Well-Being*. New York City, pp. 259–270. <https://doi.org/10.4324/9781315739793>.
- Kahneman, D. (2011) *Thinking, Fast and Slow*. London: Penguin Books.

- Kahneman, D. and Frederick, S. (2002) 'Representativeness Revisited: Attribute Substitution in Intuitive Judgment', in Gilovich, T., Griffin, D. and Kahneman, D. (eds.) *Heuristics and Biases*. Cambridge: Cambridge University Press, pp. 49–81. <https://doi.org/10.1017/CBO9780511808098.004>.
- Kelp, C. (2018) *Good Thinking: A Knowledge First Virtue Epistemology*. London: Routledge.
- Kozhevnikov, M. (2007) 'Cognitive Styles in the Context of Modern Psychology: Toward an Integrated Framework of Cognitive Style', *Psychological Bulletin*, 133(3), pp. 464–481. <https://doi.org/10.1037/0033-2909.133.3.464>.
- Kvanvig, J. L. (2003) *The Value of Knowledge and the Pursuit of Understanding (Cambridge Studies in Philosophy)*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511498909>.
- Lepock, C. (2011) 'Unifying the Intellectual Virtues', *Philosophy and Phenomenological Research*, 83(1), pp. 106–128. <https://doi.org/10.1111/j.1933-1592.2010.00425.x>.
- Lyons, J. C. (2019) 'Algorithm and Parameters: Solving the Generality Problem for Reliabilism', *Philosophical Review*, 128(4), pp. 463–509. <https://doi.org/10.1215/00318108-7697876>.
- MacIntyre, A. (1981) *After Virtue*. Notre Dame: University of Notre Dame Press. <https://doi.org/10.4324/9781912281954>.
- Mi, C. and Ryan, S. (2020) 'Skilful Reflection as a Master Virtue', *Synthese*, 197(6), pp. 2295–2308. <https://doi.org/10.1007/s11229-016-1192-z>.
- Miracchi, L. (2015) 'Competence to Know', *Philosophical Studies*, 172(1), pp. 29–56. <https://doi.org/10.1007/s11098-014-0325-9>.
- Montmarquet, J. (1987) 'Epistemic Virtue', *Mind*, 96(384), pp. 482–497.
- Montmarquet, J. (1992) 'Epistemic Virtue and Doxastic Responsibility', *American Philosophical Quarterly*, 29(4), pp. 331–341.
- Ohlhorst, J. (2022) 'Dual Processes, Dual Virtues', *Philosophical Studies*, 179(7), pp. 2237–2257. <https://doi.org/10.1007/s11098-021-01761-7>.
- Osman, M. (2004) 'An Evaluation of Dual-Process Theories of Reasoning', *Psychonomic Bulletin and Review*, 11(6), pp. 988–1010. <https://doi.org/10.3758/bf03196730>.
- Pedersen, N. J. L. L. (2017) 'Pure Epistemic Pluralism', in Coliva, A. and Pedersen, N. J. L. L. (eds.) *Epistemic Pluralism*. Cham: Springer International Publishing, pp. 47–92. [https://doi.org/10.1007/978-3-319-65460-7\\_3](https://doi.org/10.1007/978-3-319-65460-7_3).
- Plato. (2005) *Plato: Meno and Other Dialogues*. Edited by R. Waterfield. Oxford: Oxford University Press.
- Samuels, R. (2012) 'The Magical Number Two, Plus or Minus: Dual-Process Theory as a Theory of Cognitive Kinds', in Evans, J. and Frankish, K. (eds.) *In Two Minds: Dual Processes and Beyond*, pp. 129–146. <https://doi.org/10.1093/acprof:oso/9780199230167.003.0006>.
- Samuelson, P. L. and Church, I. M. (2015) 'When Cognition Turns Vicious: Heuristics and Biases in Light of Virtue Epistemology', *Philosophical Psychology*, 28(8), pp. 1095–1113. <https://doi.org/10.1080/09515089.2014.904197>.
- Sosa, E. (2015) *Judgment and Agency*. Oxford: Oxford University Press.
- Stanovich, K. E. (2009) *What Intelligence Tests Miss*. New Haven: Yale University Press.

- Stanovich, K. E. and West, R. F. (2000) 'Individual Differences in Reasoning: Implications for the Rationality Debate?', *Behavioral and Brain Sciences*, 23(5), pp. 645–665. <https://doi.org/10.1017/S0140525X00003435>.
- Sylvan, K. L. (2020) 'An Epistemic Nonconsequentialism', *Philosophical Review*, 129(1), pp. 1–51. <https://doi.org/10.1215/00318108-7890455>.
- Wright, S. (2010) 'Internalist Virtues and Knowledge', *Acta Analytica*, 25(2), pp. 119–132. <https://doi.org/10.1007/s12136-009-0066-0>.
- Zagzebski, L. (1996) *Virtues of The Mind*. Cambridge: Cambridge University Press.

# 5 Trust virtuously

## 5.1 Introduction

In Chapter 3, I argued that epistemic arbitrariness lies at the roots of the problem of demarcation, the threat of relativism, and the issue of blind trust. I defined epistemic arbitrariness as follows:

- (A) A belief that *P* is *arbitrary* if and only if there is no criterion independent from whether *P* that makes *P* preferable to some incompatible alternative *Q*.

In this chapter, I will introduce a virtue-based solution to the problem of demarcation for epistemic entitlement. I have argued that in every belief system there are fundamental hinges that we have to trust to be the case. We have to trust them to be true because any belief system necessarily has to take some things for granted. Any evidence for a hinge would already presuppose that the hinge is indeed the case.

We are entitled to trust in rule and cornerstone propositions because they enable us to be cognitively active. Without trust in some propositions, we would be epistemically paralysed (Wright, 2004, p. 191) – but truth-directed cognitive activity is an epistemic good.

However, because some instances of epistemic activity seem problematic, this leads to the problem of demarcation. And if we cannot give any reason for why our cognitive activity is preferable to some alternative (truth-directed) activity, then it seems to lack warrant. This would lead to a pernicious relativism, where any cognitive activity would produce entitlement. Arbitrary trust in hinges would be blind. At the very least, we would not be able to claim entitlement to trust our hinges.

I argued that the beliefs and belief sets of agents cannot help us to solve this problem. We should instead examine the agent. The idea is to solve the problem of demarcation using the notion of virtues (*cf.* Wright, 2004, pp. 210–211). I will show that epistemic virtue reduces the arbitrariness of

our cornerstone and rule certainties to a degree that solves the problem of demarcation. The justificational blindness of trust in hinges, which arises from the unavailability of evidence or reliability, is mitigated through the fact that entitlement arises from the exercise of our virtues rather than cognitive activity *simpliciter*.

First, I will begin by examining what role our virtues play for our hinges and entitlements. In a second step, I will argue that even though virtues may not directly contribute to our hinges' being true – that is virtue does not *justify* hinges – it reduces the hinges' arbitrariness because the trust is underwritten by a truth-directed disposition. This resolves the problem of demarcation, differentiating virtuous trust from non-virtuous trust. Meanwhile, entitlement plays an enabling role for Type 1 and Type 2 virtues.

Finally, I will present some challenges that can be raised for this solution. Notably, my account combines internalist and externalist elements in a manner that may make them sit uneasily together – this also connects to the sceptical problem. Further, there is the threat that entitlement will abolish itself by appealing to virtue. Finally, I will briefly consider how virtuous entitlement plays out in social contexts.

## 5.2 Virtues and hinges

### 5.2.1 A hinge-producing virtue?

First, I will show that our virtues play an important role for our hinges and hence influence our epistemic entitlement. By default, we should expect our epistemic virtues to influence our hinges given that virtues are not properties of single beliefs but rather the traits of the cogniser who also trusts in hinges. The question then is: How do our hinges relate to our virtues? More precisely, we have to ask: How do we manifest epistemic virtue when trusting in hinge certainties?

I cannot simply argue that, because an epistemic agent has some virtue, the hinges he or she possesses are also virtuous – which would arguably solve the demarcation problem. I cannot take this simple route because virtue theories are not built to account for hinge beliefs but for epistemic success with respect to regular beliefs. I need to deliver an argument for why our virtues are also useful for hinges.

This point can be illustrated nicely with reliable virtues. Suppose there was a faculty that reliably delivered true hinges. We could easily solve the problem of demarcation by arguing that we are only entitled to trust in reliably produced hinges. Indeed, it could be argued that, historically speaking, there have been two attempts to posit a reliable faculty for recognising fundamental truths.

The first theory of such a reliable faculty of insight was rationalism. Rationalists like Descartes (2013) and Leibniz (2014) argued that reason

is a reliable capacity that delivers insights into the fundamental nature of things. The light of reason, in their book, delivers true and even deductively certain fundamental truths. However, the rationalists also thought that reason could be rational and reflexive. In my account, such a rationalist virtue would belong to Type 2. In contrast to hinge epistemology, the rationalists thought that fundamental truths could be recognised as being self-evident, making trust unnecessary.

The rationalist project is generally considered to have failed (Coliva, 2015, p. 71). As I argued in Chapter 2, there is no rationalist capacity for insight for the hinge propositions that we need. I will not continue beating the long-dead horse of rationalism here. Humans do not possess a God-given rational insight into nature.

There is, however, a second virtue approach that argues that we have a reliable faculty that delivers hinge certainties. This faculty is, on this account, a sort of reason *light* that reliably delivers trust in fundamental truths without the accompanying self-evidence (*cf.* Coliva, 2010). This could be called the faculty of *common sense*. Common sense would also arguably be a Type 1 virtue, in contrast to rationalist reason.

Is common sense a reliabilist virtue? This depends on two questions: (a) is common sense *reliable*? and (b) is common sense a (Type 1) *faculty*? One philosophical position goes back to Thomas Reid (2012) and answers both questions with ‘yes’ (Woudenberg and Peels, 2020, p. 3).

Virtue reliabilism would have a very elegant response to the first question a), of whether common sense is reliable. Given that every agent has her own faculty and different individuals’ faculties have variable degrees of reliability, common sense could be claimed to be a virtue that some individuals have. Consider eyesight as a faculty: Some people have perfect vision, while I have to wear glasses to see further than a meter. Common sense too could be more reliable in some agents and less reliable in others. The reliability of common sense, however, is hotly contested (Woudenberg and Peels, 2020).

If there were such a faculty, its reliability would not extend very far. If common sense were a reliabilist faculty, we could still expect a considerable degree of convergence from this. While there is considerable convergence among human individuals on some fundamental principles which Spelke and Kinzler (2007) call our core knowledge, this core knowledge is quite weak. It forms a rudimentary framework that does not go beyond simplistic ideas about causation, substance, or agency; and some of it is simply inaccurate (Carey, 2009, p. 10).

But even if we granted that we might get some mileage out of the reliability of common sense; is common sense b) *a faculty*? This, I think, is the real crux of the matter. Reid (2012) claims that it *is* a faculty in a broad sense. Common sense is situated in our divinely created natural

constitution and is a set of *principles* that are part of human nature (Wood, 2020, pp. 72–73).

For common sense to be an epistemic virtue, it must be a *disposition*. Following my argument in the preceding chapter, our cognitive system must be disposed to produce hinge certainties. In this section, I will argue that common sense would be the *wrong kind* of disposition for forming a virtue. Namely, instead of a capacity with the structure of a *function*, it is simply a disposition to form a fixed *set* of beliefs (Coliva, 2010).

Given their fixed nature, hinge certainties do not usually change. Hence, such a purported capacity should always produce the same limited set of hinges. You might object that mathematical beliefs are also fixed in this sense and are thus like a common sense faculty given that the capacities that support our mathematical thinking always give the same response to the same question, if they are virtuous. Children, who know little mathematics as yet, may, for example, simply have memorised the multiplication tables from 1 to 10.

But precisely this shows the point I want to make: A virtuous mathematical capacity will deliver the correct response to *any question* that satisfies certain formal constraints. It is potentially able to form infinitely many beliefs, even if it cannot do so in practice. When exercising your arithmetical virtues, you learn something new. However, our common sense is more akin to having memorised the multiplications from 1 to 10. There is a finite set of hinge or common sense beliefs an agent will hold even though these can be reformulated and recombined in many ways. ‘Exercising’ common sense just tells you what you were already certain of. Consequently, common sense retrieves your hinges rather than producing them and cannot deliver an answer to novel questions.

A second reason why it is not plausible to consider common sense a real Type 1 faculty is that it lacks a domain. While Type 1 faculties have clearly specified domains, such as recognising faces or estimating distances, common sense is at best unified by a doxastic function. Namely, what unifies common sense beliefs is that they are about hinges. In contrast, Type 1 faculties are usually unified by their sharing some aetiological source, for example, visual signals of a certain shape, or memory traces of past events. The hinge function of certain propositions does not lead back to any shared source; at best, they could be unified by the absence of any such source. However, this does not alleviate the fact that common sense would have to produce hinges concerning such disparate domains as the nature of other minds and basic mechanics.

Finally, I want to raise a last point against there being a virtue that generates hinges, that is, a faculty of common sense. Namely, there simply isn’t a *there* there. Hinges are not the output of some faculty. Apart from the fact that it isn’t possible to identify such a faculty, there are alternative stories



that are more compelling. Notably, there is the account that I develop in this chapter. Instead of flowing from a faculty, hinges are the presuppositions for our faculties and they are necessary to the faculties' being epistemically efficacious. Positing a common sense faculty that delivers these hinges produces either an infinite regress or a circularity, as we would need to explain over and over how we get the hinges for trusting in our common sense.

In summary, I think that there are three reasons to think that we do not possess a hinge-producing common sense. First, other Type 1 virtues are functions whose output varies with their input. Common sense, however, has a fixed set as its output. Second, other Type 1 virtues are aetiologically and thematically unified, they are domain-specific. But common sense concerns all domains at once and is characterised more by its lacking an aetiological source. Third, there is a competing explanation for how we gain our entitlements, namely that we rely on them in using our other epistemic capacities. It is not clear in this case why we would require a dedicated extra faculty for hinges.

### 5.2.2 *Trust in hinges as reliance on virtues*

Common sense as we find it in our ordinary life hardly has the traits of a reliable faculty. It is rather a set of beliefs that we have acquired partly by virtue of how our cognition functions, but also partly from our social, cultural, and natural environment.

Given what we know of human cognition, I do not think that there can be either a Type 1 common-sense virtue, that reliably delivers the fundamental convictions that we trust in, or hinge insights from a rationalist Type 2 virtue. Hinge propositions are singular and play a very special role within a domain. The set of hinges is not unified in any way that would allow us to plausibly postulate a domain for which there is a specific reliable capacity. Our virtues do not directly produce hinges in the way that they produce regular beliefs.

Additionally, the idea of a Type 1 virtue that reliably delivers common sense truths implies that our common sense beliefs are reliabilistically justified. This would suggest there is a kind of evidence to be had for common sense beliefs just as there is a kind of evidence to be had for our perceptual beliefs. This, I argued in Chapter 2 does not work. There is no such reliable faculty for fundamental beliefs and that is why we need to trust in hinges.

Nevertheless, examining common sense is a fruitful pursuit, not because it is itself a virtue, but because of how it relates to our virtues. Namely, an aspect of common sense is the conviction that we have reliable capacities. It is commonsensical to rely on the faculties that we possess. I therefore think that common sense is a subset of our hinges (Coliva, 2010, p. 208). We consequently have to ask: How do we manifest epistemic virtue when trusting in hinge certainties?

Given their different structures, Type 1 and Type 2 virtues become manifest in different ways when we trust in a hinge. I will first explain how we manifest Type 1 virtues when trusting in hinges. Second, I will consider what role Type 2 virtues play for our hinges.

I just argued that common sense is a subset of our hinge beliefs. More specifically, common sense consists of the hinges that come with our Type 1 cognition.

Type 1 virtues reliably deliver true beliefs in response to certain input ranges. It is common sense to *trust* our Type 1 faculties – indeed, if we do not trust our senses and by extension our Type 1 capacities, then we cannot use them. Everyone has more or less the same Type 1 modules – in that sense, they are common. There are two ways that trust in our Type 1 faculties can make trust in hinge certainties manifest.

The first is that in order to trust our Type 1 capacities, we have to trust the rules that the capacities concerned are reliable. This also means that we can rely on their output beliefs. This is the bare minimum of reliance on and trust in our Type 1 cognition. Type 1 cognition is richer and requires more than relying on a set of rules.

The second way, that we trust our Type 1 cognition is our trusting in the cornerstones that come with the conceptual contents of its outputs. Our Type 1 cognition has a built-in conceptual structure as it deploys categories like *agent*, *cause*, or *number* (Carey, 2009). This conceptual structure can be expressed as a set of cornerstones that we have to trust to be the case when we rely on our Type 1 virtues. Recall Wright's argument for entitlement from conceptual schema in Chapter 2.

I believe that the hinges that come with trusting our Type 1 faculties are our *animal certainties*, as Moyal-Sharrock (2004) describes them. Exercising Type 1 cognition, which is *animal* and evolutionarily older, means trusting in certain hinges that are built into its capacities. In our example of facial recognition, this is the rule that we can trust our facial recognition capacities. If you successfully recognise Yasmin, then that person is indeed Yasmin. But additionally, facial recognition might be used to ascribe agency (Carey, 2009, p. 236), that is, it comes with the cornerstone that some things are agents and can autonomously cause events. Facial recognition also allows reading emotional states, which means it comes with the cornerstone that the owners of faces have an inner life. As mentioned in Chapter 1, this bears some similarities to Burgean entitlement insofar as my account also points to our natural capacities but it shifts the locus of entitlement from the capacities' outputs to their implicit presuppositions (*cf.* Greco, 2021).

A striking feature of Type 1 hinges, that follows from their animality and innateness, is that we cannot help but accepting them. They are part and parcel of our cognitive make-up and we cannot block their correspondent capacities from operating and producing the beliefs whose presuppositions

these Type 1 hinges are. Annalisa Coliva (2015, p. 139 ff.) makes a similar point, but hers is much broader; it includes essentially the whole set of our common sense hinges. Nevertheless, we are not forced to slavishly accept these animal Type 1 hinges because our Type 2 capacity can hedge them in and partially suspend them.

The hallmark of Type 2 cognition is that it goes beyond common sense. There is, as with Type 1 virtues, trust in the rule that we can rely on our Type 2 cognition. However, I do not think our Type 2 capacities come with built-in cornerstones as Type 1 cognition does.

Our Type 2 cognition can instead rely on entitlements of cognitive projects or rational deliberation. Type 2 cognition has the function of answering questions and solving problems – in other words, undertaking a particular cognitive project. However, as Wright argues, doing this necessarily means relying on presuppositions and amongst these are our hinges. Similarly, undertaking the cognitive project of deciding what to do, that is, deliberating rationally, requires reliance on some hinges.

Type 2 virtues also have a further function. As thinking dispositions, they monitor our other cognitive processes. That means that sometimes Type 2 cognition overrides our Type 1 outputs. It constrains the range of application of our Type 1 capacities. Implicitly, this also means that the rule proposition corresponding to such a capacity is restrained in its domain of application. For example, when faced with optical illusions like the Muller-Lyer illusion, Type 2 cognition constrains the rule that we can rely on our visual-spatial interpretations.

A further monitoring function that our Type 2 capacity has is allowing reflection on our hinge certainties. We do this as philosophers, but such reflection also happens in other sciences (Lakatos, 1976). It allows us to critically examine our preconceptions, although it is not clear whether or how we can effectively change them. I believe that this is the *motif* of Wittgenstein's *On Certainty* (1969). While we may not be able to rationally shed our certainties in the light of evidence, this sort of reflection may nevertheless recontextualise or restrain the role of a hinge certainty by constraining its scope. Given this function, Type 2 cognition also gives us a way to grapple with deep disagreement about our hinges. Through this reflective capacity, we may be able to understand the nature of deep disagreement, a possibility that we would not have without certain Type 2 virtues.

In sum, exercising Type 1 virtues implies relying on a set of built-in hinges. However, the exercise of Type 2 virtues has a more complex relation to hinges. On the one hand, the exercise of Type 2 virtues implies a reliance on hinges through cognitive projects and a reliance on our Type 2 capacity. On the other hand, Type 2 cognition allows cognitive access to our hinges – either by monitoring the functioning of other Type 1 and 2 processes or by directly examining our doxastic states, that is, reflection.<sup>1</sup>

### 5.3 Virtuous hinges and the problem of demarcation

Consequently, Type 1 and Type 2 virtues play a role for different hinges and reduce their arbitrariness in different ways. In this section, I will argue that Type 1 virtues reduce the arbitrariness of the rule and cornerstone presuppositions required for exercising Type 1 faculties, while Type 2 virtues reduce the arbitrariness of other hinges. If virtues successfully reduce our hinges' arbitrariness, then this will also solve the other accompanying issues, namely relativism and deep disagreement, as well as bizarreness. Recall how I defined arbitrariness:

- (A) A belief that P is *arbitrary* if and only if there is no criterion independent from whether P that makes P preferable to some incompatible alternative Q.

Arbitrariness undermines our claim to epistemic warrant, by rendering the epistemic standing of incompatible hinges indistinguishable. This leads to problematic forms of relativism and opens the gate to bizarre hinges.

#### 5.3.1 *Virtuous Type 1 hinges*

Type 1 virtue is enabled by entitlement to trust in certain hinges because the hinges play a presupposition role for the virtue. For cognisers to be able to possess Type 1 virtues, they require certain hinges that allow the Type 1 faculties to become epistemically efficacious. Without those presupposition hinges, the cogniser would be unable to form her beliefs virtuously. Hence, our Type 1 virtues transcendently warrant their own presupposed hinges.

As belief-forming capacities, our Type 1 faculties have two kinds of presuppositions: First, the rules indicating that we can rely on what this faculty delivers and second, the cornerstones for trusting in the basic ontology that our faculties deliver.

The first way of reducing arbitrariness is our having to presuppose a rule proposition in order for a Type 1 faculty to be a virtue. If we do not implicitly trust in a faculty's correspondent rule, it remains epistemically inert. This is very similar to Burgean entitlement as it also points to our natural reliable faculties. However, my account of entitlement is narrower because it only points to our Type 1 virtues. Additionally, entitlement is not assigned to every single Type 1 belief, but rather to the virtuous agent's implicit presuppositions that go into possessing that virtue.<sup>2</sup>

The second way follows entitlement from our cognitive schema. Our Type 1 cognition comes with certain ontological categories, it puts things in terms of causes, agents, objects, etc. We would be unable to epistemically rely on our Type 1 faculties if we didn't trust them to represent the world as it is. Thus, possessing a Type 1 virtue also entitles us to trust some cornerstones.

The demarcation of hinges deriving from the Type 1 faculties with which we find ourselves consequently depends on whether the faculties are in fact virtues or not. This distinguishes trust in the hinges of Type 1 virtues from hinges that do not enable any virtue to function. We are only entitled to trust the former, thus resolving the arbitrariness problem because this makes virtuous hinges preferable to non-virtuous hinges.

This account of entitlement to trust in our Type 1 virtues is a neo-Reidian reintroduction of common sense. Trusting and relying on our Type 1 virtues as well as their presuppositions is commonsensical. However, I do not make a bare appeal to common sense because my account also makes a naturalist appeal to our Type 1 faculties. In this sense, my account of Type 1 entitlement is close to that of Burge.

As mentioned, relativism is a very limited issue here, given that we are all roughly the same organisms who share the same cognitive resources (*cf.* Coliva, 2015, p. 145). Additionally, if two agents' distinct Type 1 virtues happened to have incompatible hinge presuppositions, this seems to be a harmless degree of arbitrariness and hence of relativism. After all, both could point to a virtue that is, respectively, enabled by these hinges.

Finally, this also gives us a way to deal with a supposed entitlement to rely on bizarre or extraordinary faculties as the faculty has to be a virtue in order to generate entitlement. As contrast cases, consider the hypothetical extraordinary Type 1 faculty of being able to tell immediately whether a number between 1 and 1,000,000 is prime and also the hypothetical extraordinary faculty of being able to tell how another person is feeling without even seeing or hearing them, that is, telepathy (Coliva, 2015, p. 140).

On my account, there are clear tools to determine whether these two faculties are actually Type 1 virtues. We can (empirically) examine whether the subjects actually possess such a Type 1 disposition as well as whether it is indeed reliable and a virtue. The easiest way is investigating whether other people have the same faculty. Hence, our naturalism about virtues gives us the tools for examining whether a hinge is arbitrary or not and, consequently, whether the agent is entitled to trust her faculties or not. This verification would be a Type 2 process, as is clear. But note that, given entitlement, we are not required to go through with this verification.

Another example of an extraordinary faculty would be a *sensus divinitatis* that some people claim to possess, that is, a supposedly reliable faculty that is capable of delivering accurate religious beliefs (Plantinga, 2000). Given its quasi-perceptual nature, it would arguably be a Type 1 faculty. We could thus examine it and see whether people do indeed possess such a faculty and whether it delivers stable outputs. Stability of outputs is necessary for it to be a stable and a reliable disposition, given the supposed unchanging nature of the object of this faculty. Notably, Type 1 virtues are extremely stable in their outputs across different cultures. Humans from

Fireland to Sakhalin have the same basic perception of causality or agency, for example. This is not the case for a supposed *sensus divinitatis* which is usually attached to the locally prevalent religious framework. Hence, claiming entitlement for one's religious hinges, especially cornerstones, on the basis of one's supposed *sensus divinitatis* does not hold up. Additionally, it could be debunked as an unreliable manifestation of our Type 1 faculty for detecting agency.

### 5.3.2 *Virtuous Type 2 hinges*

Wright formulated the problem of demarcation in terms of 'responsible belief management' (Wright, 2014, p. 245). Reliance on arbitrary hinges is epistemically irresponsible. Type 2 virtues are the responsibilist virtues; they require a stable desire for truth to be stably responsibilist. Consequently, an agent possessing Type 2 virtues essentially exemplifies cognitive activity aimed at truth and respect for truth. If possessing a Type 2 virtue means that one manifests the quality that is the source of epistemic entitlement, then this removes arbitrariness. If, as a virtuous agent, my cognitive activity is stably aimed at truth, then the presuppositions required for said activity are not arbitrary.

But how exactly does Type 2 virtue deal with arbitrariness of hinges, and what sort of entitlement can we expect from this? Given that Type 2 virtue is complementary with Type 1 virtue, the corresponding entitlements are also complementary. That is, our entitlement deriving from Type 2 virtue expands beyond common sense and constrains the reach of common sense where necessary. If our entitlements were constrained to common sense, we would be epistemically limited to a common sense empiricism. We could only rely on our Type 1 faculties and their accompanying common sense ontology.

As I argued in Chapter 4, Type 2 virtues can realise epistemic value in many ways. Consequently, Type 2 virtues can reduce the arbitrariness of our hinges in many ways. Entitlement of cognitive project is one way of obtaining entitlement over and beyond our common sense hinges. The first way that our Type 2 virtues reduce our hinges' arbitrariness and give you a claim to entitlements is by guiding cognitive projects. The second way is through the monitoring function of our Type 2 cognition which reduces the arbitrariness of the hinges enabling other virtues and thereby grants you *a claim* to entitlement to trust these other hinges.

Recall that entitlement of cognitive project derives its warrant from the fact that an investigation cannot get off the ground without our taking some presuppositions as our starting point. However, if we were to try to undertake any arbitrary cognitive project, our correspondent entitlement would also be arbitrary which would undermine our claim to it.

Consequently, if Type 2 virtue is to reduce the arbitrariness of the entitlement of a cognitive project, then Type 2 virtue has to constrain epistemic agents in the sort of investigations they undertake. Given my account of the nature of Type 2 virtues, I believe this is eminently plausible.

Type 2 virtues allow agents to successfully solve problems. This involves accurately recognising whether a given question is a real problem that does in fact need to be solved. Type 2 virtues also include competence in assessing the appropriateness of a given approach to the problem (Sosa, 2015, pp. 51–52).

To show the contrast, an agent who possesses Type 2 virtues will thus undertake different cognitive projects to an agent who lacks such virtues. Type 2 virtue is geared towards producing diverse epistemic values: Truth, understanding, and justification. Hence, if an agent with such virtues undertakes such a cognitive project, this cognitive project will also be geared towards these values. This will involve cognitive activity aimed at truth, more precisely aimed at truth under a wide range of circumstances, because Type 2 virtues are *stable* dispositions. Consequently, the presupposed hinges will also play the role of permitting such a stably excellent investigation – this generates the responsibilist entitlement to trust in the cognitive project's hinges. Consider, for example, the years of strenuous and diligent work by Marie Skłodowska-Curie. She maintained her research project throughout many years and faced all sorts of obstacles – even if her project had failed, she was entitled to trust in her research's presuppositions, and she could claim this entitlement.

However, also an agent who lacks these Type 2 virtues may start out with a cognitive project that aims at discovering the truth. But because she lacks the relevant stable dispositions, she could lose track of her project more easily. She could also become satisfied too quickly with her results. As an example for this, consider Montmarquet's (1992) study of Hitler which characterises him as a close-minded epistemic agent who took all data to confirm his views. In summary, a non-virtuous agent's cognitive projects are only aimed at truth contingently, if at all. Consequently, the correspondent hinges only contingently permit cognitive activity aimed at truth. This does not suffice to overcome the arbitrariness of our hinges; hence we are not entitled to trust in our non-virtuous cognitive projects.

However, this does not mean that we are only entitled to the presuppositions of a cognitive project when we possess fully developed Type 2 virtues. An agent's Type 2 virtues may be more or less developed thus allowing them to execute more or less demanding cognitive projects. I can gain entitlement for the presuppositions required in finding out how much I weigh but I would lack the necessary entitlements for finding out on my own how much I would weigh inside a black hole. In the former case, I possess the necessary Type 2 capacities, but in the latter, I fall significantly short of the required virtues.

The second way that Type 2 virtue reduces the arbitrariness of our hinges is by monitoring other cognitive processes. Type 2 virtues monitor both Type 1 and Type 2 processes. This function may be compared to Mi and Ryan's (2020) virtue of skilful reflection, but I assign it exclusively to a Type 2 capacity. I will argue that this monitoring function reduces the arbitrariness of hinges by rendering us accountable for and giving us a sensitivity to the functioning of Type 1 and 2 capacities corresponding to the hinges.

In the first case, Type 2 monitoring ensures that Type 1 faculties only run in contexts where they are in fact reliable. This is what Samuelson and Church's (2015) proposed virtue of intellectual humility does. This reduces the arbitrariness of entitlements by restricting the breadth of the hinges. It constrains the rule proposition corresponding to the Type 1 capacity to those contexts in which the capacity is in fact reliable. In this way, Type 2 monitoring virtues give us a claim to our entitlement to trust in our Type 1 capacities. This is also important because this monitoring helps us to recognise non-normal environments and this recognition also can override our reliance on our Type 1 faculties, that is, it restrains the scope of a Type 1 rule.

Humans have the ability to manipulate their environment in ways to systematically trick Type 1 cognition so that Type 2 monitoring is required. As a simple illustration of this, consider a house of mirrors. Our Type 1 representation of space and the location of objects no longer functions reliably in a house of mirrors – you are unable to accurately locate objects through Type 1 cognition. Our Type 2 monitoring virtues are able to recognise this breakdown and suspend assenting to our Type 1 outputs. Thus, Type 2 virtues can effectively constrain the scope of our trust in the rule that our spatial perception is reliable. We bracket this Type 1 rule in a house of mirrors.

Note that Type 2 monitoring of our Type 1 faculties does not necessarily make the latter more reliable. That is, I do not think that such monitoring reduces a hinge's arbitrariness by making the corresponding processes more reliable. A first, very straightforward way in which monitoring reduces the arbitrariness of relying on Type 1 faculties is by making the process more *sensitive* to mistakes. This is what Samuelson and Church (2015) mean by epistemic humility. Monitoring makes a trusted rule proposition less prone to produce mistakes, for example, unwarranted inferences, because it makes the corresponding faculty less prone to doing so.

The second way that monitoring reduces arbitrariness is that it generates accountability. Type 2 monitoring, being controlled and active, produces a track record of what an automatic and unconscious Type 1 capacity does and does not do. When relying on a Type 2-monitored Type 1 capacity rule or cornerstone presupposition, you can better understand what you effectively rely on. This reflective access also differentiates monitored cognitive processes from unmonitored ones. The latter simply produces cognitive outputs that an agent relies on. I would argue that rules and cornerstones about



capacities are less arbitrary if agents are more sensitive to the mistakes these capacities can make and if agents can account for the underlying processes. This access explains why we can claim entitlement for monitored processes.

However, our Type 2 virtues can also monitor our Type 2 processes to reinforce our claim to their corresponding entitlements. This monitoring then feeds back into the cognitive projects that an agent undertakes. Through our Type 2 monitoring, the arbitrariness of our cognitive projects can be further reduced. This also enables the agent to better evaluate whether they are indeed able to carry out the project.

For example, consider a philosopher undertaking a research project about virtues. Her monitoring Type 2 virtues continuously track the status and progress of this project. This reduces the arbitrariness of the project because it is under continuous control which makes the agent more sensitive to mistakes and accountable regarding what she is doing. Consequently, the hinges presupposed in the project are also less arbitrary. This shows that arbitrariness of a cognitive project is not only dependent on the project itself but also on the agent carrying it out.

Consider again the contrast between an unmonitored and a monitored Type 2 cognitive activity. In the former case, the agent executes the project just in the way that it comes to their mind, without a plan for, an overview of, or any reflection on what is being done. The absence of active monitoring raises the risk of unnoticed mistakes and the agent cannot account for their project as a whole. In contrast to this, a monitored cognitive activity is more sensitive to mistakes because monitoring also involves checking. Additionally, the agent knows what they have been doing and what they still need to do to successfully carry out the project; that means they can give an account of it. The hinge presuppositions for the latter cognitive project are arguably less arbitrary than the hinge presuppositions for the former because the agent is more sensitive and accountable in the latter case. The agent virtuously monitoring also is aware of their hinge presuppositions, and the enabling role they play for the project. This gives them a claim to being entitled to trust these hinges.

Philosophical cognitive projects are especially peculiar because it is in the nature of philosophical investigation that any presupposition can be up for debate. Hence, the threat of arbitrariness is particularly acute for philosophical cognitive projects. Diligent monitoring of how the project evolves may reduce the arbitrariness insofar as the investigation is then undertaken virtuously.

The foregoing account does not completely eliminate relativism concerning entitlement of cognitive project. There will still be incompatible virtuous cognitive projects that generate claims to incompatible entitlements. However, this class of potential incompatible entitlements is much smaller than the class of *all* entitlements flowing from any cognitive project aimed at truth (Hazlett, 2014).

I do not think that a weak relativism is problematic *per se* if it is constrained to virtuous cognitive projects. When we look at real instances of valuable epistemic practices, then communities carry out different and incompatible cognitive projects in parallel. These may have incompatible presuppositions but demanding coherence from society seems to be an unhelpful constraint. Philosophical research is a telling illustration of this phenomenon. Philosophical cognitive projects run the gamut of possible presuppositions: Some presuppose idealism, others materialism, and others still pragmatism. In ethics, we find everything from presupposing moral realism to nihilism. Essentially, almost any research project may be virtuously investigated within philosophy.

Virtuous agents carrying out these incompatible cognitive projects can safely bracket the incompatible entitlements of others. That is, it is right for them to be certain of their hinges because the cognitive projects they are pursuing would otherwise be impossible – this is their claim to entitlement. The virtue requirement also gives us a criterion for demarcating which incompatible entitlements epistemic agents should at least acknowledge and which they can safely ignore. Non-virtuous cognitive projects do not need to be taken seriously. Hence, when a philosopher writes about logics and assumes bivalence, she should at least acknowledge that Graham Priest denies it, but she does not have to deal with the denials of bivalence that she could find in her local bookstore’s ‘metaphysics’ section if it is assumed that the former is a virtuous epistemic agent while the latter authors are not.

This brings me to the last feature of the problem of demarcation: Bizarreness. As I argued, bizarreness is itself hinge-relative and therefore not a useful marker. But epistemically virtuous incompatible beliefs must be taken seriously even though they appear bizarre. I recall, for example, a discussion between a philosopher and a quantum physicist where the philosopher refused to even consider the possibility that quantum states did not follow bivalence. The physicist’s theory was bizarre to the philosopher as quantum physics was bizarre to Einstein. However, assuming the physicist to be exhibiting the necessary Type 2 virtues, the philosopher should be aware of and contend with the physicist’s incompatible hinges even though they appear bizarre.

### 5.3.3 *Competition: other virtue hinge epistemologies*

Interestingly, John Greco (2021) and Ernest Sosa (2021) have recently developed similar arguments that hinge epistemology and virtue epistemology are intimately connected. However, given their virtue epistemological background, they come from the converse direction than me. They take virtue epistemology as their point of departure and introduce hinge epistemology as a solution to the former’s problems instead of departing from

hinge epistemology and going to virtue. This leads to both commonalities and differences in our outlooks.

Greco takes an economical perspective on epistemology: Epistemic goods like economic goods can be produced and transmitted and virtue plays a key role for these activities. Hinges, on this picture, are like common or public goods – everybody can use them for free – they are *common knowledge*<sup>3</sup> (Greco, 2021, p. 110).

Greco (2021, p. 124) justifies that we know our hinges in the same way that I argue that we are entitled to trust our hinges: Accepting hinges *constitutively enables* us to exercise our virtues. Differently from Greco, I do not think that this transcendental argument gives us enough epistemic force to *know* rather than *warrantedly trust* our hinges.

Sosa's view is more complex: Like Greco, he argues in a neo-Moorean fashion that we *know* most hinges (Sosa, 2021, p. 212). We know these hinges through the intuitions that are competently produced by a competence or virtue (Sosa, 2021, p. 216). I have argued earlier that such common sense or rationalist capacities, and hence knowledge of these hinges, are not possible.

However, Sosa thinks that there is a special class of hinges which he calls a competence's *background conditions* which seem to coincide with a virtue's normalcy conditions (Sosa, 2021, p. 127). According to Sosa, we cannot believe or know whether they hold; instead we blamelessly – but not warrantedly – *assume* them.<sup>4</sup> This is in line with Pritchard's (2016) and Coliva's (2015) rejection of entitlement and their rejection of the idea that hinges are beliefs.

I think that entitlement to trust is the more appropriate epistemic state both with hinges in general and our virtues' background conditions in particular. This is because blameless assumption is doxastically too flimsy; we can blamelessly assume almost anything almost always, but we will lose such an assumption at the slightest epistemic resistance or counter-evidence. For instance, we can assume a proposition to be true in order to refute it in a *reductio ad absurdum*. In particular, we would shed such an assumption if we were *credibly* informed of the likely threat of a sceptical scenario (*cf.* Sosa, 2021, p. 129).

## 5.4 Challenges for virtuous trust

### 5.4.1 Externalism through the back door?

We now have to deal with the pressing objection that my account of Type 1 virtue is reliabilist and therefore externalist. Entitlement is supposed to deliver an internalist solution to the problem of demarcation – to underwrite an *internalist* warrant to claim entitlement. Externalist solutions for scepticism are not satisfactory to the internalist.

There would be a simple solution to this problem: If we were able to claim entitlement because we knew that we were virtuous, then we would simply have to appeal to the fact that we possess the said virtues to gain entitlement to trust in its corresponding hinges. Knowledge of virtue would furnish all we needed.

I have argued earlier that we do not have any such knowledge independently of our already trusting our virtues to work. Also, Linda Zagzebski (2009, pp. 77–78) argues that we need to trust our virtues, even though we cannot know that they are virtues. I consequently do not think that we need to know that we possess the corresponding virtues in order to have entitlement to trust in our hinges. Instead, we need to be aware of our virtues' operation in order for our entitlements to not be arbitrary. We need a simple grasp of our virtues' operation, and this simple grasp ensures that our reliance on our virtues is not completely blind.

By simple grasp, I do not mean a full-blown theory of epistemic virtue such as I presented in Chapter 4. Sosa (2021, p. 219) also argues that we do not need to have a full-blown theory of a competence in order to rely on it. We instead need to be aware that we possess perceptual and reasoning faculties that can work better or worse depending on context – these are our common sense Type 1 virtues. Imagine the theoretically least demanding conception of epistemic virtue. Call this notion *simple virtue*. For example, children understand from early infancy that they can see and that other people also can see, and that almost everyone has these virtues to a certain extent.

Being able to grasp such simple virtue is required for *simple non-arbitrary entitlement*. The access to a simple notion of virtue is the internalist element required to have epistemic entitlement from a virtuous Type 1 cognitive activity because this reliance is not blind. Hence, you may be utterly mistaken about, or lack any awareness of, the mechanisms behind these virtues, but you are at least aware of their operations in yourself and others. You have to appeal to this operation to claim simple entitlement to trust in your capacities. Norman the clairvoyant, who just happens to find himself saddled with certain reliably formed beliefs but isn't aware of his faculty's operation could therefore not claim entitlement for the rule that beliefs delivered by his clairvoyance (of which he lacks any notion) are reliable (Bonjour, 1980).

Your entitlement to trust in a virtue's presuppositions would be lost if you did not have this simple grasp of its operation. This lack would mean that you are completely unaware of this faculty, that is, that this faculty's operation happens to you in the same way as to a biological automaton. Take, for example, the right-hand bias that most humans have, their unconscious preference for objects on their right over objects on their left. Most of us are completely unaware of this bias. It may even be reliable in some (social) contexts because we share it with other humans who also

tend to put important things to their right. However, we are not entitled to rely on it given that we have no awareness whatsoever of its operation.<sup>5</sup>

As a consequence, animals with no cognitive access to their capacities (worms, for example) also lack entitlement to rely on these capacities. Meanwhile, rabbits are arguably aware that they see and thus they are entitled to rely on their vision. This is a departure from Crispin Wright's (2014, p. 243) view that there is no entitlement at the level of such animal hinges.

This answer does not yet deal with the sceptical problem about whether or not you can know that your capacities are indeed virtuous. To demand this degree of self-knowledge is to ask for too much. As I just argued, doing away with the arbitrariness of our entitlement requires less than this.

First, we do know that everyone possesses these faculties – everybody has more or less the same Type 1 capacities that are similarly virtuous and presuppose the same hinges. Additionally, we cannot help but trust our Type 1 capacities as a matter of biological fact.

Second, I think the responsibilist monitoring function of our Type 2 virtues makes monitored virtues preferable to those that run uncontrolled or tend to be overridden. Notably, the monitoring grants us a *claim to entitlement* to trust in hinges. Hence, we do not trust our capacities arbitrarily if we trust responsibly on the basis of our monitoring capacity. However, this *claim to entitlement for responsible trust* is already more sophisticated than simple entitlement to trust in our simple virtues. Consequently, there are also two types of entitlement, *simple* and *responsible*. Note that there is also a claim to responsible entitlement that is not built upon simple entitlements: Responsible entitlement of cognitive project enabling your exercise of Type 2 virtues. Wright (2014, p. 243) limits entitlement to these claims, that is, to the level of Type 2-monitored reflective knowledge.

I argued that even with simple entitlement there is a minimal awareness of a virtue's operation, thus making it minimally internalist. Unfortunately, not all our capacities are virtues, even though we are aware of them; that is we do not know whether we are simply entitled to trust these capacities. Consider a rabbit's flight response. Clearly, the rabbit is somehow aware of this response and relies on it but the response itself is not virtuous because it produces many false positives.

Table 5.1 Entitlements for the types of virtues

	<i>Type 2</i>	<i>No Type 2</i>
<b>Type 1</b>	Monitored responsible claim to entitlement	Simple entitlement
<b>No Type 1</b>	Responsible claim to entitlement of cognitive project	–

Why is the rabbit not entitled to rely on that faculty even though entitlement is supposed to be internalist and the rabbit is aware of its operation and it cannot help but trust it? Because the rabbit could not gain responsible entitlement for this response if it had the prerequisite capacities, that is, Type 2 cognition. Were it to monitor its flight response virtuously it would stop trusting it.

Consider a similar phenomenon in humans: Our innate fear of spiders. Given that this fear of spiders is unreliable, we are not entitled to rely on it in all environments. Indeed, by exercising the Type 2 capacity of reflecting on whether and which spiders are actually dangerous, the reliance on this unreliable Type 1 capacity can be overridden. There is no responsible entitlement to be had for this capacity in many environments and thus there isn't any simple entitlement to be had for it either. Consequently, simple entitlement is only possible in cases where the agent could gain a responsible claim to entitlement if they had the prerequisite Type 2 monitoring virtues.

#### 5.4.2 *Scepticism dispelled?*

While I did not set up this work as an argument against scepticism, it obviously does have something to say about scepticism. After all, this book is an account of epistemic entitlement and entitlement is supposed to be a response to the sceptical challenge.

Essentially, my response to the sceptical challenge follows Wright's solution: The entitlement theorist concedes to the sceptic that our hinges are not justified and that we do not know them. It therefore is a sceptical response to the sceptical problem (Wright, 2004, p. 206). Nevertheless, we are warranted to trust hinges to be true because we are entitled by their enabling our cognitive activity aimed at truth and the exercise of our virtues which manifests our respect for the truth.

For Wright, the whole sceptical debate plays out at the level of our *claim to knowledge* or what Sosa would call *reflective knowledge* (Wright, 2014, p. 243). By introducing also the simple entitlement of a cogniser being aware of their virtue's output, I have also expanded entitlement to the animal level, although not all animal knowledge is supported by entitlement because it can be underwritten by blind reliance on a capacity without even awareness of the capacity.

It is no problem that this pre-reflective, unentitled level of cognition offers no solution to the sceptical problem. To a flatworm, there is no epistemic difference whether it is in a sceptical scenario because it lacks even the most basic awareness of its capacities' operation or of an external world beyond its stimuli. Type 2 virtue is the source of the entitlement that overcomes the sceptical problem, but it is also the faculty that first generates the sceptical problem (Wright, 2014, p. 243).

When you are only capable of trusting in simple virtues, that is, to have simple entitlement, then the sceptical problem does not arise for you, doxastically speaking, whether you are a flatworm or a rabbit. To *understand* the sceptical problem, you first need to possess a well-developed Type 2 cognition and use it to monitor your own cognition with it. This means the capacity that makes the sceptical problem – your lacking claim to justification – accessible to you is also the capacity that allows you to trust in your virtues responsibly. To that end, you need to develop your Type 2 virtues and monitor your epistemic activity. Without responsible entitlement, there is no sceptical paradox and *vice versa*.

You might now worry about how this deals with the new evil demon problem (Lehrer, 1983). Is a brain in a vat, with exactly the same epistemic Type 1 and Type 2 virtues as me, exactly as entitled and justified as me? I would say that yes, this envatted brain can warrantably undertake the same cognitive activities and trust their presuppositions to be true just as I can. In other words, my account remains true to its internalist colours notwithstanding the introduction of the externalist virtue motif.<sup>6</sup>

But would the possession of Type 1 virtues not be undermined, if a brain has been envatted from the start?<sup>7</sup> Such a brain's Type 1 virtues never responded to any real environmental inputs towards which reliabilist Type 1 virtues are geared. Even a brain envatted for all of its life would be a human brain which possesses human Type 1 cognition. Consequently, the brain can still exercise this Type 1 cognition virtuously or not.<sup>8</sup>

Would the envatted brain's life-long predicament not undermine its being Type 1 virtuous however? This brain never has been reliable because it always was in an epistemically hostile environment – how could it then be virtuous? And if my envatted-brain counterpart isn't virtuous, how could I be entitled to trust that I have Type 1 virtues? Would this trust not be arbitrary? Indeed, an envatted brain would never be able to exercise its Type 1 virtues successfully.

Still, the Type 1 capacities' exercise would have been reliable under normal circumstances if the brain had had a body. Virtue is dispositional reliability, not reliability across the board. Type 1 cognition, and consequently entitlement to trust it, does not need to be sensitive to envatment scenarios in order to be virtuous. Otherwise, no brain – whether embodied or envatted – would possess Type 1 virtues and be entitled to trust those. This follows Burge's (2003) strategy about entitlement to trust in normal environments.

The importance of Type 2 virtues for scepticism points to another issue that I raised in Chapter 3. Could not a virtuous epistemic agent claim entitlement of cognitive activity for her sceptical investigations? Why not be a sceptic or anti-realist and claim entitlement for this?<sup>9</sup>

While I do not think that what I have said about epistemic Type 2 virtues up to now gives us an argument against someone claiming entitlement for their

sceptical investigations, I do think that an argument against sceptical cognitive activity can be made. Namely, I would argue that fully endorsing scepticism is a kind of epistemic defeatism, an intellectual vice that is the contrary of investigative curiosity, and this vice undermines the sceptic's virtues.

Let me explain: By endorsing global sceptical hinges, that is, the certainty that we lack cognitive access to the world, we capitulate in face of the epistemological difficulties that arise from sceptical arguments and their ilk. We give up the hope of finding out anything that goes beyond our small Cartesian theatre. This is also a kind of intellectual cowardice: Type 2 monitoring serves to avoid errors, and scepticism is the ultimate error-avoidance strategy, just as doing nothing is the ultimate harm-avoidance strategy. But such cowardice overlooks the benefits of more ambitious cognitive activity. For that reason, I believe that sceptical cognitive projects are not virtuous, and there is no virtuous epistemic agent undertaking a sceptical cognitive project.<sup>10</sup>

Additionally, the hinges which we would trust in a global sceptical investigation are incompatible with the hinges that we are entitled to trust given our Type 1 and Type 2 virtues. By trusting global sceptical hinges, we would contradict our hinge certainties that enable us to trust and exercise our Type 1 and Type 2 capacities. If we did not trust these hinges, we could not be cognitively active – we would be cognitively paralysed. Thus, sceptical investigations even keep us from being virtuous.

#### 5.4.3 *Virtuous astrology?*

I argued earlier that bizarreness and relativism are sufficiently curtailed by the requirement for virtue. However, you might object that, even if we require virtues, could there not be problematic virtuous cognitive projects? Can you really have an entitlement to trust in astrology as long as you are virtuous?

Michael Lynch (2012) argued that this would lead to far-reaching practical problems. Namely, democratic and civil societies would be unable to come to a common understanding prerequisite for large-scale coordination. He proposes the solution that we should seek a collective reflective equilibrium through an epistemic veil of ignorance and a procedure to agree on a generally acceptable common epistemic denominator. Lynch calls this procedure the *method game* which would arguably support epistemic virtues. Unfortunately, the method game does not work. The basic problem is that incompatible cognitive projects would be preserved through the proposed procedure, consequently also bizarre cognitive projects would not be cancelled (Kappel, 2012; Ohlhorst, 2021).

Consider, for example, the fact that Isaac Newton was an avid alchemist and Tycho Brahe was an astrologer. Both indubitably were intellectually



virtuous. Consequently, according to my view, they were entitled to trust some strange hinges, and I embrace that conclusion.

However, I want to argue that nowadays it would be much harder to be an intellectually virtuous astrologer. In order to count as virtuous, an epistemic agent has to be aware of and sensitive to what his or her epistemic peers and authorities know. Not only do they need to be sensitive to it, they have to take an interest in the current state of knowledge. They have to understand what the received opinion is and *why* it is the received opinion. Only then can they undertake a virtuous cognitive project. This point harks back to one of the earliest contributions to responsibilist virtue epistemology: James Montmarquet's (1992) arguments showing that simple sensitivity to available evidence is insufficient.

Astrological research would be subject to severe constraints given contemporary knowledge about astronomy and empirical research methods. Thus, a virtuous epistemic agent might become an astrologer. However, it would be difficult for her because there is no theory available concerning astrology's putative mechanisms. Additionally, human affairs, the subject of astrology, are extremely messy and a virtuous epistemic agent would be aware that this messiness makes genuine and precise astrological predictions almost impossible. For these reasons, there are arguably no entitlements to trust in astrology, because the corresponding cognitive projects are not feasible and a virtuous epistemic agent would not undertake them.

This argumentative strategy has some similarities to Briesen's (2012) argument that bizarre cognitive projects are blocked by the results of preceding cognitive projects, but it filters this consideration through the virtuous agent and a society's broader cognitive achievements. That is, rather than just appealing to whatever her epistemic community already has established through past cognitive projects, the virtuous epistemic agent is also sensitive to considerations of feasibility, expectations of success, and disagreeing opinions by other virtuous epistemic agents.

We have to consider an issue related to bizarre but virtuous astrology: There can also be *morally bad* hinges. Take, for example, the dogmatic racist's certainty that people of certain ethnicities are inherently inferior. Could the racist not be epistemically entitled to trust this bad hinge if she is epistemically virtuous? Could there not be racist cognitive projects that give her entitlement for her morally bad hinge? This would be a bad result.<sup>11</sup>

Montmarquet (1992) and Hazlett (2014, p. 7) address exactly this issue. The former requires conscientiousness and a love of truth for the agent's good epistemic standing; the latter requires that we not be *epistemically vicious* in our cognitive activity. But the example can be constructed in a way that we assume that the racist is epistemically virtuous but has a racist hinge.

I can only sketch a solution here: The relevant entitlement is for Type 2 cognitive projects. In Chapter 4, I suggested that our Type 2 virtues cannot

be properly separated from our moral virtues. The problem of morally bad hinges shows to which extent this is true: Cognitive activity cannot only be epistemically arbitrary but also morally.

Following Hazlett (2014), I think we need an anti-vice requirement – but in my case, it is against *moral* vices: If a cognitive project enables the exercise of a moral vice, then this taints or threatens the entitlement. Even if we assume that the psychologists developing the CIA's ‘enhanced interrogation’ torture programme were epistemically virtuous, they were not morally entitled to trust their presuppositions because their presuppositions were tainted by their moral vices.

An alternative possible response is to argue that the epistemic and the moral virtues are the same and that the dispositions aimed at truth, that is, epistemic virtues, coincide with dispositions aimed at the good, that is, moral virtues. This can be derived from the (controversial) traditional Aristotelian doctrine that all (moral) virtues are one virtue – to possess one virtue, you need to acquire them all (Toner, 2014). Zagzebski (1996) argues that if the moral virtues are unified in this way, then plausibly also the epistemic virtues belong in the bundle. In that case, the psychologists’ moral viciousness would extend to their epistemic capacities.

#### 5.4.4 *Eliminativism about entitlement*

Nikolaj Pedersen (2020, p. 315 ff.) has developed a challenge to consequentialist accounts of entitlement. If trust in hinges is instrumental to producing true beliefs, then what role does entitlement as a peculiar kind of warrant play? All we need is the consequentialist derivative value of how trust produces true beliefs in the (expected) long run. On this view, entitlement is not a distinctive kind of warrant, rather it is justification from expected value.

My account, while transcendental, also introduces an independent kind of epistemic value to solve the problem of demarcation: Virtue. Why should virtuous trust not be warranted simply because of its virtuousness? Does entitlement really do anything when we appeal to epistemic virtue?

I believe that this argument would gain considerable traction if my account argued (like Sosa) that our entitlements are a *product* of our virtuous cognitive activity. Because, in that case, the hinges would derive their value from the underlying virtue, just as chairs made by an excellent carpenter will derive their value from excellent carpentry and tend to be excellent chairs.

However, I explicitly argued that virtuous entitlement does not arise as a product of virtuous Type 1 or Type 2 cognitive activity. Instead, it permits it. It is the necessary precondition for being cognitively active. The entitlements play an enabling role for the virtues rather than being their outputs.

Using our example of the carpenter, the entitlements are more akin to the carpenter's tools. The carpenter is entitled to use the tools because she is a good or virtuous carpenter. A bad carpenter would exhibit bad workmanship even if he happened to produce a good chair and he is therefore less entitled to use the tools than the skilled carpenter. Analogously, a virtuous cogniser is entitled to trust her hinges because this enables her to exercise her virtues in contrast to the vicious cogniser.

The carpenter's tools are not a product of the capacity but a precondition for the carpenter's exercise of her excellent carpentering. This is the value that entitlement enables; the exercise of an excellent capacity that manifests good craftsmanship. Usually, the carpenter trusts her tools to work, but this is also the point at which the analogy with trust in hinges breaks down. There are circumstances at which the carpenter would check and examine whether her tools are still working, and she would be able to do so – we do not and cannot verify our hinges.

Because the epistemic virtues depend on entitlements in order to be exercised, epistemic entitlement cannot be reduced to the virtues' values. However, the virtues cannot be reduced to entitlement either.

#### 5.4.5 *The social nature of virtuous entitlement*

The last question I want to examine has been mentioned before. Namely, the preceding account is very demanding because it requires virtuous entitlement. Are we only entitled to trust in hinges that are presuppositions related to virtues that we actually possess? My notion of Type 2 virtues is quite demanding – are then most of us limited to simple entitlement to trust in our Type 1 virtues?

I for one lack the prerequisite capacities and virtues to think about and investigate the nature of special or general relativity. Consequently, on my current view, I am unable to claim entitlement to trust in any hinges about the relative nature of space and time. If you look at new-age caricatures of physical theory then these constraints are somewhat plausible. However, I don't have any strange new-age theories about space-time. Is any scientific knowledge forbidden epistemic fruit for the layperson who lacks the prerequisite virtues? I think we can avoid this unfavourable conclusion.

First, note that everyone has a certain degree of virtuous Type 2 capacities and is thus entitled within the possibilities of their virtues. However, this still prevents us from attaining difficult knowledge that we could not acquire on our own.

Recall Zagzebski's (1996, pp. 280–281) argument that I can exhibit a virtue, without possessing it by acting like a virtuous counterpart. In that case, I could gain the corresponding entitlements if I went through the calculations on the relativity of space-time under instruction from a physicist,

much as students do in tutorials. Because I would be behaving as my counterpart who possesses the prerequisite virtues and capacities, I would be entitled to trust this particular cognitive project's presuppositions.

I think Zagzebski's account gets it almost right. However, virtue and entitlement do not come from my behaving like a virtuous counterpart. Instead, I rely on my instructor's virtue and entitlement to gain my own.

Namely, epistemic virtue can be socially distributed from virtuous experts to laypeople who lack the prerequisite virtues. I believe that we can rely on others' virtue by being entitled to trust in these virtuous experts' testimony. That expert testimony will function like a hinge for me. Recall Greco's (2021) common knowledge: we need some shared epistemic resources that we can all draw on in order to be epistemically successful.

This entitlement requires not only virtue of the expert but also my own virtue. This is a form of entitlement of cognitive project. I can learn about the world from an expert, but I need to exercise certain Type 2 virtues to actually gain an understanding of what is being explained to me. If I fail, I might form strange new-age beliefs about relativity theory. I also need a certain degree of virtue to recognise good experts. In sum, I can claim entitlement to trust in expert testimony if I trust responsibly and exhibit certain Type 2 virtues.

This shows that we can defer to other subjects' virtues. Indirectly, this means that we also defer to their hinges. This follows roughly Tyler Burge's (1979) conception of how we often defer to experts in our use of concepts. Consequently, I may have some hinges of the form *X, whatever the epistemically virtuous experts trust it to be*.

Inspired by Burge's approach, Sanford Goldberg (2010) has developed an anti-individualist extended epistemology. Goldberg's account is reliabilist, but it shows how an individual's good epistemic standing depends in many ways on their epistemic community. We epistemically rely on each other not only in cases of testimony but also when we defer to experts about background theories. My account is a virtue- and hinge-epistemological version of this anti-individualist model. I have not given many details of how this works, but the key mechanism is our relying on others' epistemic virtues and therefore, indirectly, on their hinges.

## 5.5 Responsible trust

In this chapter, I argued that epistemic virtue reduces the arbitrariness of our trust in hinges to a degree that solves the problem of demarcation as well as its associated problems. First, I showed how Type 1 and Type 2 virtues relate to trust in hinges, especially because they rely on hinges to function. Second, I showed how the virtues reduce the arbitrariness of our trust and solve the problem of demarcation by constraining the

scope of some hinges and by allowing virtuous cognitive projects. Third, I responded to some objections to this account, namely, externalism, scepticism, eliminativism, and the fact that Type 2 virtuous are too demanding. I also extended entitlement into the social sphere.

## Notes

1. Gerken (2020) has an interesting alternative account of entitlement and our cognitive capacities. Modifying Burge (2003), entitlement does not apply to hinges but to beliefs produced by perceptual processes; meanwhile, justification is generated by the exercise of our faculty of reason. This could easily be formulated in terms of Type 1 and Type 2: We are entitled to believe in the outputs of our Type 1 capacities, while Type 2 cognition generates justification. Note, however, that this account, like Burge's, omits the presupposition role played by hinge certainties.
2. A further hinge that we arguably rely on for our Type 1 faculties is that we are in a sufficiently normal environment for the faculty to function reliably. As I mentioned in Chapter 4, humans have transformed their environment to a degree that we cannot always take normalcy for granted. I will argue in Section 5.3.2 that Type 2 therefore plays an important role for human Type 1 faculties (*cf.* Sosa, 2021).
3. This metaphor raises an interesting question about whether there is an epistemic tragedy of the commons? For instance, whether there is a sort of epistemic pollution? *Prima facie*, I suspect that the problem of demarcation points in this direction: Tainted hinges destroy and undermine our epistemic commons.
4. Sosa (2021, pp. 220–221) needs these blamelessly assumed background conditions because he sees them as the only way out in the case of a combination of lottery scenarios and Cartesian scepticism where you would with high probability have been in the bad case but end up in the good case.
5. Contrast this with the availability heuristic, with whose operation we are arguably acquainted through feelings of familiarity. However, there may be many Type 1 processes that we are not entitled to rely on because they fly entirely under our radar. Note, however, that I claim that blocking entitlement requires a strong sense of inaccessibility.
6. Sosa, who is committed to *knowledge* of our hinges takes a markedly different turn: He would deny that the envatted brain is a person and can have virtues (Sosa, 2021, p. 176). I do not know what goes into personhood and what does not, however, I would deny that envatted brains could not have virtues. I defined virtues as epistemically excellent dispositions of Type 1 and 2 cognition – also an envatted brain will have such dispositions. Unfortunately, however, the brain is in a very unusual scenario, and it cannot exercise its virtues – an envatted brain knowing of its predicament could explore its condition and successfully exercise epistemic virtues. For a brain to be in the same situation as an embodied counterpart, it cannot simply be fed inputs, the simulation also needs to respond to the brain's outputs.
7. I would like to thank an anonymous referee for raising this issue.
8. I am excluding functionally equivalent swamp brains that have been created in freak coincidences.
9. Annalisa Coliva (2015, p. 145) argues that this means opting out of the practice of epistemic rationality. I opt for a different though similar strategy. I do

- not think that scepticism, etc. are arational as Coliva suggests, just as liar-paradoxes aren't asemanitic. They teach us something about the structures and limits of systems like language and rationality.
10. See also Wright (2014, p. 244). I have to add an important *caveat* here: There are cognitive projects that look very similar to a sceptical cognitive project, but which do not arise out of that vice. Namely, I am thinking of the Cartesian (Descartes, 2013), Kantian (Kant, 1998) or Husserlian (Husserl, 1913) cognitive projects. Each of those cognitive projects has some kind of sceptical presupposition as its starting point, the epochè as Husserl calls it. These projects, I believe, are not sceptical investigations, but methodological scepticisms that use their sceptical hinges to bracket confounding factors. Scepticism in this case functions like the laboratory or idealised model does for the natural scientist.
  11. I thank an anonymous referee for raising this point.

## References

- Bonjour, L. (1980) 'Externalist Theories of Empirical Knowledge', *Midwest Studies in Philosophy*, 5(1), pp. 53–74. <https://doi.org/10.1111/j.1475-4975.1980.tb00396.x>.
- Briesen, J. (2012) *Skeptische Paradoxa*. Paderborn: Mentis.
- Burge, T. (1979) 'Individualism and the Mental', *Midwest Studies in Philosophy*, 4(1), pp. 73–121. <https://doi.org/10.1111/j.1475-4975.1979.tb00374.x>.
- Burge, T. (2003) 'Perceptual Entitlement', *Philosophy and Phenomenological Research*, 67(3), pp. 503–548. <https://doi.org/10.1111/j.1933-1592.2003.tb00307.x>.
- Carey, S. (2009) *The Origin of Concepts*. Oxford: Oxford University Press.
- Coliva, A. (2010) *Moore and Wittgenstein – Scepticism, Certainty, and Common Sense*. New York City: Palgrave Macmillan.
- Coliva, A. (2015) *Extended Rationality*. Basingstoke: Palgrave Macmillan.
- Descartes, R. (2013) *Meditations on First Philosophy*. Edited by J. Cottingham. Cambridge: Cambridge University Press.
- Gerken, M. (2020) 'Epistemic Entitlement – Its Scope and Limits', in Graham, P. J. and Pedersen, N. J. L. L. (eds.) *Epistemic Entitlement*. Oxford: Oxford University Press, pp. 151–178. <https://doi.org/10.1093/oso/9780198713524.003.0004>.
- Goldberg, S. C. (2010) *Relying on Others*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199593248.001.0001>.
- Greco, J. (2021) *The Transmission of Knowledge*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108560818>.
- Hazlett, A. (2014) 'Entitlement and Mutually Recognized Reasonable Disagreement', *Episteme*, 11(1), pp. 1–25. <https://doi.org/10.1017/epi.2013.46>.
- Husserl, E. (1913) *Logische Untersuchungen, Erster Band: Prolegomena Zur Reinen Logik*. Halle an der Saale: Max Niemeyer.
- Kant, I. (1998) *Critique of Pure Reason*. Edited by P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.
- Kappel, K. (2012) 'The Problem of Deep Disagreement', *Discipline Filosofiche*, 22(2), pp. 7–25. <https://doi.org/10.1400/205566>.
- Lakatos, I. (1976) *Proofs and Refutations*. Cambridge: Cambridge University Press.
- Lehrer, K. (1983) 'Belief, Acceptance and Cognition', in Parret, H. (ed.) *On Believing. Epistemological and Semiotic Approaches*. Berlin: Walter De Gruyter, pp. 172–183. <https://doi.org/10.1515/9783111727967-012>.

- Leibniz, G. W. (2014) 'La monadologie/Monadologie', in *Monadologie und andere metaphysische Schriften*. Hamburg: Meiner, pp. 110–151.
- Lynch, M. P. (2012) *In Praise of Reason*. Cambridge, MA: MIT Press.
- Mi, C. and Ryan, S. (2020) 'Skilful Reflection as a Master Virtue', *Synthese*, 197(6), pp. 2295–2308. <https://doi.org/10.1007/s11229-016-1192-z>.
- Montmarquet, J. (1992) 'Epistemic Virtue and Doxastic Responsibility', *American Philosophical Quarterly*, 29(4), pp. 331–341.
- Moyal-Sharrock, D. (2004) *Understanding Wittgenstein's on Certainty*. Basingstoke: Palgrave Macmillan.
- Ohlhorst, J. (2021) 'Epistemic Austerity: Limits to Entitlement', *Synthese*, 199(5–6), pp. 13771–13787. <https://doi.org/10.1007/s11229-021-03397-w>.
- Pedersen, N. J. L. L. (2020) 'Pluralist Consequentialist Anti-Scepticism', in Graham, P. J. and Pedersen, N. J. L. L. (eds.) *Epistemic Entitlement*. Oxford: Oxford University Press, pp. 297–326. <https://doi.org/10.1093/oso/9780198713524.003.0011>.
- Plantinga, A. (2000) *Warranted Christian Belief*. Oxford: Oxford University Press.
- Pritchard, D. (2016) *Epistemic Angst*. Princeton: Princeton University Press.
- Reid, T. (2012) *An Inquiry Into the Human Mind*. Cambridge: Cambridge University Press.
- Samuelson, P. L. and Church, I. M. (2015) 'When Cognition Turns Vicious: Heuristics and Biases in Light of Virtue Epistemology', *Philosophical Psychology*, 28(8), pp. 1095–1113. <https://doi.org/10.1080/09515089.2014.904197>.
- Sosa, E. (2015) *Judgment and Agency*. Oxford: Oxford University Press.
- Sosa, E. (2021) *Epistemic Explanations*. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198856467.001.0001>.
- Spelke, E. S. and Kinzler, K. D. (2007) 'Core Knowledge', *Developmental Science*, 10(1), pp. 89–96. <https://doi.org/10.1111/j.1467-7687.2007.00569.x>.
- Toner, C. (2014) 'The Full Unity of the Virtues', *The Journal of Ethics*, 18(3), pp. 207–227. <https://doi.org/10.1007/s10892-014-9165-2>.
- Wittgenstein, L. (1969) *Über Gewissheit – On Certainty*. Edited by G. E. M. Anscombe and G. H. von Wright. Oxford: Blackwell.
- Wood, P. (2020) 'The Scottish School of Common-Sense Philosophy', in Peels, R. and Woudenberg R. V. (eds.) *The Cambridge Companion to Common Sense*. Cambridge: Cambridge University Press, pp. 66–85. <https://doi.org/10.1017/9781108598163.004>.
- Woudenberg, R. V. and Peels, R. (2020) 'Introduction : Why Common Sense Matters', in Peels, R. and Woudenberg R. V. (eds.) *The Cambridge Companion to Common Sense*. Cambridge: Cambridge University Press, pp. 1–16. <https://doi.org/10.1017/9781108598163.001>.
- Wright, C. (2004) 'I – Crispin Wright: Warrant for Nothing (and Foundations for Free)?', *Aristotelian Society Supplementary Volume*, 78(1), pp. 167–212. <https://doi.org/10.1111/j.0309-7013.2004.00121.x>.
- Wright, C. (2014) 'On Epistemic Entitlement (II)', in Dodd, D. and Zardini, E. (eds.) *Scepticism and Perceptual Justification*. Oxford: Oxford University Press, pp. 213–247. <https://doi.org/10.1093/acprof:oso/9780199658343.003.0011>.
- Zagzebski, L. (1996) *Virtues of the Mind*. Cambridge: Cambridge University Press.
- Zagzebski, L. (2009) *On Epistemology*. Belmont: Wadsworth.

# Conclusion

In this work, I defended the view that we are entitled to trust in our hinge certainties. Hinges are fundamental propositions that we have to presuppose in order to epistemically engage with the world. This grants entitlement, a form of epistemic warrant. I primarily dealt with the objection that hinge certainties are arbitrary and that we therefore cannot possess any warrant for them. The solution to this arbitrariness problem is that we are only entitled to trust in hinge certainties that enable the activity of our epistemic virtues. For this purpose, I developed a special view of virtues that appeals to dual process theory.

The concepts of certainty, hinge, and trust are fundamental for our epistemology. They explain how our most fundamental convictions are structured and what role they play for our ordinary beliefs. Some propositions possess the status of hinges in a belief set because they play an implicit presupposition role for all other beliefs. There are two ways of playing that role: First, a hinge may be a cornerstone that is entailed by the other beliefs and whose acceptance is therefore a necessary precondition for rationally holding these other beliefs. Second, a hinge may be a rule proposition to which the agent is committed due to the (rule-bound) ways in which she changes her beliefs.

We have to be certain of these hinge commitments. Certainty here means doxastic certainty in the sense that we cannot change these commitments rationally. Doxastic certainty does not only occur with hinges; we are certain of a great many things. If our hinges were subject to revision, our entire belief systems would undergo continuous Copernican shifts. Additionally, hinges dictate the way we revise our beliefs therefore they cannot be rationally revised themselves.

Hinge certainties are, however, not acquired like other beliefs. We instead trust them to be true. Due to their pre-epistemic hinge role, we cannot gain regular non-circular justification for these hinge certainties. This is due to the fact that hinges determine what counts as evidence for what. Hinges, *qua* certainties, would simply be evidence for themselves, and all other



evidence that could speak in their favour presupposes these hinges. For that reason, we cannot but trust in such propositions.

Nonetheless, trust in hinges is not removed from all epistemic evaluations. While we cannot be justified foundationally or coherentistically to trust in our hinges, we are nevertheless *entitled* to trust in hinges. Namely, we are transcendently entitled to trust our hinge certainties to be true, because this enables cognitive activity that is aimed at truth.

This contrasts with consequentialist accounts of entitlement that argue that we are warranted to trust in our hinges because the expected epistemic outcome is better if we trust in our presuppositions than if we do not. On my view, we are entitled to trust in our rules and cornerstones because they are a necessary precondition for the possibility of cognitive activity aimed at truth and such cognitive activity is an inherent epistemic good.

Entitlement of cognitive activity is an internalist kind of warrant and thus it has to contend with the problem of the criterion (Chisholm, 1973), which Wright (2014) calls the demarcation problem. The problem boils down to the fact that different epistemic activities cannot be distinguished with respect to their doxastic structure. Neither hinge is preferable to the other. This arbitrariness threatens epistemic warrant.

Additionally, this problem relates to further epistemological difficulties, namely those of pernicious relativism and bizarre hinges that nevertheless would seem to gain entitlement. Any two instances of earnest cognitive activity may generate entitlement, even strange cognitive projects. Consequently, unrestricted entitlement of cognitive activity aimed at truth does not work, because the corresponding hinges are arbitrary. Nothing distinguishes one hinge from the other epistemically; this undermines claim to warrant and hence entitlement.

To deal with this problem, I proposed a novel account of epistemic virtue. There are two disparate accounts of virtue, reliabilism and responsibilism, that have been taken to be incompatible. I argued instead that they are complementary because they are the virtues of two complementary information processing types of our cognitive apparatus. These two types are described in dual process theory (Evans and Stanovich, 2013).

Reliabilist virtues are virtues of our Type 1 cognition, which is characterised by its fast and automatic information processing. This fits the characteristics of our reliabilist virtues which also do not require conscious control. Meanwhile, our responsibilist virtues are virtues of our Type 2 cognition, which is sequential and controlled. Again, epistemic agency and control is key to responsibilist virtue. Consequently, reliabilist and responsibilist virtue are related to each other analogously to how our two cognitive types are related: They are complementary. My approach also outlines a clear naturalisation programme for virtue epistemology, designating the cognitive structures required for possessing an epistemic virtue.

This account of epistemic virtue ameliorates my theory of entitlement of cognitive activity and solves the problem of arbitrariness. Bare cognitive activity aimed at truth can be too arbitrary, therefore entitlement is limited to virtuous cognitive activity. We are entitled to trust in hinge certainties that are a precondition for virtuous cognitive activity. Consequently, our entitlement to trust in hinge certainties arises from the fact that they transcendently enable the exercise of our virtues.

This account of non-evidential epistemology and epistemic virtue has many interesting consequences. I will just mention a few here. First, it brings epistemic internalism and externalism together: While these are independent as forms of justification, they are related at the level of entitlement. Simple entitlement to trust in our Type 1 virtues is externalist, while claim to responsible entitlement arising from our Type 2 virtues is internalist. Consequently, the two forms of warrant will be complementary. Both kinds of entitlement are, however, subject to the internalist issue of arbitrariness and require a basic awareness of our virtues' operation. Therefore, Type 2 virtues play an important role in monitoring Type 1 capacities so as to reduce their arbitrariness. This has some similarities to Greco's (2010) account of internalist and externalist justification.

Further, we can rely on others' virtues and their corresponding entitlements. This means that entitlement can be socially distributed and is not individualist. Such an individualism is a *prima facie* weakness of standard virtue theories, because virtues are usually an individual agent's virtues. Entitlement theories also have such an individualist bent, given their internalism, which is itself based on the agent's individual epistemic activity.

Finally, this proposes a mitigated relativist solution to the demarcation problem and relativism (Laudan, 1983; Boghossian, 2006). My account does not subscribe to semantic relativism but it concedes a limited epistemic relativism. Namely, the incompatible presuppositions of two incompatible cognitive projects may have the same degree of warrant – that is, both agents may claim entitlement to trust these presuppositions – as long as both agents manifest virtue in doing so. This is permissible, because a virtuous epistemic agent will keep track of the successes and failures of alternative incompatible investigations and thereby of their presupposed hinges. Keeping track of potential alternatives is part and parcel of Type 2 virtue.

This point is particularly well-illustrated in philosophical investigation. This is due to the fact that philosophical research very often is about our hinges. If philosophers are examining hinges and taking divergent approaches, they forcibly will disagree about their research projects' implicit presuppositions. But this does not undermine their philosophical projects as long as the philosophers concerned manifest epistemic virtue in pursuing said projects.

## References

- Boghossian, P. (2006) 'What Is Relativism?', in Boghossian, P. (ed.) *Truth and Realism*. Oxford: Oxford University Press, pp. 13–37. <https://doi.org/10.1093/acprof:oso/9780199288878.003.0002>.
- Chisholm, R. M. (1973) 'The Problem of the Criterion', in *The Foundations of Knowing*. Minneapolis: University of Minnesota Press, pp. 61–75.
- Evans, J. S. B. T. and Stanovich, K. E. (2013) 'Dual-Process Theories of Higher Cognition: Advancing the Debate', *Perspectives on Psychological Science*, 8(3), pp. 223–241. <https://doi.org/10.1177/1745691612460685>.
- Greco, J. (2010) *Achieving Knowledge*. Cambridge: Cambridge University Press.
- Laudan, L. (1983) 'The Demise of the Demarcation Problem', in Cohen, R. S. and Laudan, L. (eds.) *Physics Philosophy, and Psychoanalysis – Essays in Honor of Adolf Grünbaum*. Dordrecht: D. Reidel Publishing Company, pp. 111–128.
- Wright, C. (2014) 'On Epistemic Entitlement (II)', in Dodd, D. and Zardini, E. (eds.) *Scepticism and Perceptual Justification*. Oxford: Oxford University Press, pp. 213–247. <https://doi.org/10.1093/acprof:oso/9780199658343.003.0011>.

# Index

- acceptance: and belief 5–6, 21, 30, 55–58; and entitlement 7, 45–46, 50, 52, 55–58; and hinges 7–11, 24–26, 86, 119–120, 128; and rationality 59–60; and scepticism 16–18; and trust 26–28
- achievement: cognitive 45, 49–51, 75, 94, 96–97, 134; epistemic 36, 45, 60–61, 94, 97–98, 105; *see also* epistemic success
- acting *see* agency
- agency 9–11, 24, 57–58, 116, 119, 123; and virtue 87, 94, 105
- agnosticism 81
- alchemy 50, 133
- Anscombe, G. E. M. 11
- a priori 39–40, 41, 50
- arbitrariness 53, 74–75, 79–83, 85–86, 96, 106; and demarcation 86–87, 114–115, 121–126, 129–130, 135, 141–143
- Aristotle 91, 94–95, 135
- assumption 19, 38, 48–49, 75, 77, 128
- astrology 64, 84, 133–134
- awareness 27, 101, 126, 129–131, 134
- axiom 40, 42, 65
- belief: ordinary 6, 12, 30, 48–49, 82, 141; perceptual 17, 43–44, 61, 67, 118; religious 6, 18, 122–123; systems 5–8, 11, 14, 23, 74, 78; true 36, 51–53, 63–64, 78, 87, 104–106; *see also* acceptance; doxastic state; rationality and belief
- bizarreness 74–78, 83–86, 121–122, 127, 133–134
- Block, N. 11
- Boghossian, P. 78, 143
- bootstrapping 13; *see also* circularity
- Briesen, J. 54, 75, 83–86, 134
- Burge, T. 40, 97, 137; on entitlement 60–62, 67, 119, 121–122, 132
- capacities 40–41, 79, 93, 98–99, 116–118, 135–137; Type 1 101–103, 119–121, 125–126, 129–133; Type 2 104–109, 120, 124, 131–133; *see also* innate capacities; reasoning
- Carey, S. 10, 13, 98, 119; on core cognition 20–22, 25, 42, 101, 116
- Carroll, L. 9, 9n5
- certain *see* certainty
- certainty 5–6, 14–23; conditional/unconditional 14–16, 22, 30, 50; and entitlement 35, 44–46, 55, 62, 76–78, 134; hinge 13–14, 23–29, 30, 37–38, 115–120, 141–143; objective 12; psychological 14, 16, 37; and trust 26–29, 37, 50, 118–120, 133
- Chalmers, D. 3, 11, 64
- circularity 12–13, 27–29, 36–41, 43–44, 49–50, 81; non- 13, 46, 59, 61, 73, 76
- Clifford, W. K. 52, 105
- cognition: core 20–25, 43–44, 56, 62, 65, 101; Type 1 98–103, 105,

- 107, 119–121, 125, 132; Type  
2 98–100, 102–109, 120, 123,  
131–132; *see also* capacity;  
knowledge, core
- cognitive activity 55–56, 62–67, 81–82,  
126, 133–135, 142–143; and  
virtue 114–115, 123–124, 129,  
131–135; *see also* achievement;  
entitlement
- cognitive projects 83–86, 105, 126–127,  
133–138; and virtue 91, 96, 120,  
123–124; *see also* entitlement;  
inquiry; investigation
- cognitive style 105
- coherentism 13, 44–45, 74, 79, 82, 87
- Coliva, A. 49, 58–60, 63–66, 79,  
116–120, 122; on entitlement  
36, 55, 60, 86, 128; on rules  
6–7, 9–10, 21–22
- commitment 6, 8–11, 21, 40, 43–44, 47
- common sense 13, 79, 116–120,  
122–123, 129
- community 9, 12, 37, 127, 134, 137
- competence 22, 90–91, 94, 108–109,  
124, 128–129; *see also* Sosa, E.
- complementariness 91, 97, 100,  
123, 142
- conceptual role semantics 11–13,  
20–21, 24
- conciliationism 19, 37
- conditions on entitlement: negative 36,  
50–51, 62, 73–75, 78–82, 86;  
positive 36, 50–51, 62, 75, 79
- confidence 5, 15–16, 19, 23–24, 27
- consequentialism 36, 47, 51–55,  
63–67, 79, 135
- content *see* conceptual role semantics;  
cornerstone
- cornerstones 23–28, 38–42, 47–56, 74,  
119–123; and content 5–10,  
12, 23, 25, 40, 119; *see also*  
hinge; rule
- Davies, M. 24, 50
- defeat 14–15, 23–24, 29, 80, 85, 105;  
of entitlement 50, 62, 73, 75,  
81–82, 87
- delusion *see* pathological
- demarcation problem 73–78, 82–87,  
114–115, 121–123, 127, 137;  
*see also* Wright, C.
- Descartes, R. 17, 40, 47–48, 48n14,  
52, 115
- disagreement 20, 64–65, 134; deep 12,  
18–20, 37, 65, 120–121
- disposition 21, 90, 104–109, 115, 122,  
135; and character 87, 95, 98,  
124; cognitive 91–98, 100–101,  
117, 120
- dogmatism 42–43, 44n11
- doubt 14–15, 24, 47, 54, 60; *see also*  
Wittgenstein
- doxastic state 5–6, 26–27, 29, 35, 58,  
81; *see also* belief
- Dretske, F. 12
- dual process theory 91, 98–101, 99
- entailment 7–12, 23, 39–40, 48, 61,  
82–83
- entitlement: claim to 75, 81–84, 87,  
122–127, 128–132, 136–138;  
of cognitive activity 55–58,  
62–67, 73–76, 80–82, 114–  
115, 131–135; of cognitive  
project 53–55, 57–58, 75–76,  
83–86, 123–127, 130–139;  
externalist 60–61, 66, 67,  
128, 143; simple 129–132,  
130, 136, 143; *see also*  
conditions on entitlement;  
consequentialism
- epistemic agent 87, 90, 115, 124, 127,  
132–134
- epistemic alchemy *see* alchemy
- epistemic community *see* community
- epistemic goods 44, 50, 58, 105, 114,  
128; truth as an 52, 63–65, 92;  
*see also* epistemic value
- epistemic success 2, 47, 64, 104, 115;  
*see also* achievement
- epistemic value 44, 55, 58, 104–106,  
123–124, 135; truth as an 52,  
63–64, 92–93, 101; *see also*  
epistemic good
- evidence: counter- 15, 24, 62, 73, 85,  
128; insensitive to 6, 15, 22–24,  
73; self- 40, 116
- externalism 41, 60–61, 81, 115, 128,  
132; *see also* justification;  
warrant
- faculty 121–123, 125, 131; perceptual  
61, 93, 129; reliabilist 41–42,  
90, 93–94, 101, 108–109,  
115–118; *see also* reliabilism
- Foley, J. 5n1, 58

- foundationalism 13, 28, 37, 42–43  
 framework 12, 35, 53, 56, 116, 123
- generality problem 41n7, 93, 106  
 generosity of entitlement 76, 81;  
*see also* demarcation problem
- Goldberg, S. 7, 137  
 Goldman, A. 2, 19, 92, 97  
 Greco, J. 10, 22, 119, 127–128, 137,  
 143; on virtue 91, 94, 104, 106
- habit 91, 93, 95–97, 105–106  
 Hazlett, A. 18–19, 63, 101, 134–135;  
 on entitlement 36, 78, 126  
 hinge: animal 25, 119–120, 130;  
 proposition 5–11, 23–27, 30,  
 35–37, 45–50, 118; *see also*  
 certainty; cornerstone; rule;  
 Wittgenstein L.
- Hume, D. 17–18, 47–48  
 hyperintellectualise *see* intellectualise
- implicit 26, 42, 56, 62, 77, 99; certainty  
 16, 21; premise 10, 12–13, 21,  
 36, 40, 43–44; presupposition  
 54, 119, 121, 141, 143  
 indubitability 19, 28  
 infinite regress 13, 38n3, 54, 118  
 infinitism 13  
 innate 20–25, 37, 42, 56, 97, 131;  
 capacities 93, 119  
 inquiry 13, 37, 47, 54; *see also*  
 cognitive project; investigation
- Inquisition, Spanish *see* unexpected
- intellectualise 61, 91, 96–97  
 internalism 28, 35–36, 58–59, 81, 115,  
 128–132; *see also* justification;  
 warrant
- intuition 39–41, 43, 79, 98, 100, 128  
 investigation 13, 53–55, 63, 66–67,  
 75–76, 80–84; and virtue 91,  
 96, 123–124, 126–127,  
 132–133, 136; *see also*  
 cognitive project; inquiry
- irrationality 9, 14, 27, 53, 59, 84–85;  
*see also* rationality
- James, W. 27n21, 52  
 justification 12–13, 17–18, 21, 35–36,  
 45–46, 63; evidential 28, 37–39,  
 47–50, 60–61, 80; externalist  
 41, 143; immediate 42–44;  
 internalist 39–40, 47, 82–83,  
 104; reliabilist 37, 39–43,  
 86, 118; *see also* a priori;  
 coherentism; evidence; warrant
- Kahneman, D. 92, 98–99  
 Kant, I. 7, 55, 63–64  
 knowledge: animal 131; core 20–22,  
 116; first 41n6, 93n3; procedural  
 22; reflective 130–131
- Kusch, M. 78, 80n8
- leaching 82, 87  
 Lynch, M. 13, 19, 37–38, 133
- mathematics 15–16, 40, 77, 79, 117  
 monitoring 120, 123, 125–126,  
 130–133, 130
- Montmarquet, J. 87, 94–95, 104–105,  
 124, 134  
 Moore, G. E. 18, 48, 50, 128  
 morality 87, 90–91, 95, 104, 127,  
 134–135  
 motivation 87, 91, 95–97, 106;  
*see also* virtue
- Moyal-Sharrock, D. 6, 26–28, 42,  
 65, 74; on animal/objective  
 certainties 12, 21–22, 25, 119
- Oblomovian challenge 60, 65  
 over-intellectualise *see* intellectualise
- pathological 6, 12, 23, 28, 65  
 Pedersen, N. J. L. 87, 97, 105; on  
 entitlement 36, 52–53, 62, 79,  
 81, 135
- perception 20, 25, 57, 79, 105, 125; of  
 causality 22, 122–123; *see also*  
 belief; faculty
- philosophy 15, 64–65, 120, 126–127  
 Plato 29, 95n5  
 pluralism 64, 91, 105  
 Popper, K. 74  
 possibility *see* possible world  
 possible world 11, 14, 48, 56, 76,  
 108–109  
 pragmatism 18, 26–30, 80  
 pragmatism 51, 57–58, 67, 86, 127  
 presupposition *see* implicit,  
 presupposition
- Pritchard, D. 18, 22, 49, 51, 63, 74;  
 on belief and commitment  
 6, 128
- Pryor, J. 42–43

- Quine, W. V. O. 7–8, 40n4
- rationalism 39, 41, 115–116, 118, 128
- rationality 19, 51, 73, 84–86, 104; and belief 7–9, 12, 14–15, 53, 75; deliberation 57–58, 63, 66, 67, 120; epistemic 14, 59–60, 65, 67; extended 58–60, 66, 67; practical 67; *see also* reasons; intuition
- reasonableness 29, 52, 58, 74, 80, 86
- reasoning 90, 98–100, 106, 109, 115–116, 129; *see also* capacities; competence
- reasons 14, 17, 65, 67, 76, 114; and certainties 5, 24–25; epistemic 19, 21, 28, 37, 58–62, 81; and hinges 7, 10, 39, 73, 83–84; pragmatic 28–29, 51; *see also* evidence; pragmatism
- reflection 37, 102, 120, 125–126, 130–131, 133
- Reid, T. 13, 18, 79, 116
- relativism 78–79, 81–86, 114, 121–122, 126–127; epistemic 11, 45, 73–75, 90, 133; semantic 11
- reliabilism 13, 41, 82, 137; agent 91, 94, 106–109; virtue 90–93, 96–97, 101, 105, 116, 128; *see also* justification; virtue
- reliability 7, 13, 40–45, 61–62, 115–116, 129; trigger 102–103, 107, 109; unreliability 43, 52, 102–103, 107, 109, 131; *see also* reliabilism; virtue
- responsibility 27, 74–75, 82, 123, 130–132, 137; *see also* virtue
- rules 5–7, 9–13, 23–24, 38–41, 65, 93; and cognition 21–25, 43–44, 101, 103, 119–121, 125; and entitlement 55, 74, 119–121, 125, 129; and experience 25, 43–44, 48, 79; and rationality 59–60; *see also* hinge; Coliva
- sceptic 16–18, 40, 51, 59–60, 64, 131–133; anti- 13, 35, 43, 46, 56–57; argument 24, 37, 47–49, 66; investigation 76, 132–133
- scepticism *see* sceptic
- self-deception 76
- semantic *see* conceptual role semantics
- Sosa, E. 13, 50, 127–129, 128n4, 131, 135; on virtue 91–94, 106, 108–109, 124, 132n6
- Stanovich, K. 92, 99, 104–105
- Sylvan, K. 63–65, 92, 104
- transcendental 53–56, 61–62, 67, 73, 121, 135; argument 36, 47, 58–59, 65, 128
- trust: blind 27–28, 35, 114–115, 129, 131; epistemic 5–6, 26–29, 35–36, 45, 74; pragmatic 26–29, 30; social 27, 29, 29n24, 136–137; *see also* certainty
- truth: aiming at 45, 76, 81, 87, 123–126, 135; love of 64, 91–92, 95–96, 134; respect for 63–67, 73, 92, 104–105, 123, 131; *see also* truth-directed
- truth-directed 26–27, 62–64, 66–67, 87, 114–115; *see also* truth
- Type 1, 2 *see* cognition; virtue
- underdetermination 47, 80–81
- understanding 10–12, 39–40, 76–78, 94, 132–134, 137; the value of 60, 64, 104–106, 124
- unexpected *see* Inquisition
- Van Cleve, J. 13, 43
- veritism 63–64, 92–93
- vice 95, 133–136
- virtue: auxiliary 93, 107–108; reliabilist 90–94, 96–97, 100–103, 106–109, 116, 128; responsibilist 91–92, 94–97, 100, 103–109, 123, 134; simple 129–132; Type 1 100–104, 116, 118–123, 128–129, 132, 136; Type 2 103–109, 115, 118–121, 123–127, 130–134, 136–137; *see also* Sosa, E.
- Vogel, J. 13, 38, 80
- warrant: externalist 3, 60–61, 67; internalist 3, 47, 49, 73, 81, 128; and justification 36, 43–50, 60–61, 66, 86–87, 135; *see also* entitlement; justification

Wittgenstein, L. 11–16, 27, 38n1, 53,  
77, 120; on doubt 24, 47, 60;  
on hinges 6–7, 22  
world view 6, 26, 40, 77–79,  
82, 86  
Wright, C. 6, 8, 10, 17, 35–36, 87;  
on demarcation 73–76, 78,

81–82, 84–85, 114, 123; on  
entitlement 47–50, 51–62,  
64–67, 119–120, 130–131; on  
trust 26–27, 35–36, 114  
Zagzebski, L. 64, 95, 105, 107, 129,  
135–137