

Previous philosophy titles published in the Palgrave Handbooks series:

The Palgrave Handbook of German Idealism Edited by Matthew C. Altman

Forthcoming philosophy titles in the Palgrave Handbooks series:

The Palgrave Kant Handbook Edited by Matthew Altman

The Palgrave Handbook of Practical Animal Ethics Edited by Andrew Linzey and Clair Linzey

The Palgrave Handbook of the Afterlife Edited by Yujin Nagasawa and Benjamin Matheson

The Palgrave Handbook of Leninist Political Philosophy Edited by Tom Rockmore and Norman Levine

The Palgrave Handbook of the Philosophy of Aging Edited by Geoffrey Scarre

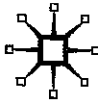
The Palgrave Handbook of Philosophy and Literature Edited by Barry Stocker and Michael Mack

The Palgrave Handbook of Philosophical Methods

Edited by

Chris Daly
University of Manchester, UK

palgrave
macmillan



Introduction, selection and editorial matter © Chris Daly 2015
Individual chapters © Contributors 2015

All rights reserved. No reproduction, copy or transmission of this publication may be made without written permission.

No portion of this publication may be reproduced, copied or transmitted save with written permission or in accordance with the provisions of the Copyright, Designs and Patents Act 1988, or under the terms of any licence permitting limited copying issued by the Copyright Licensing Agency, Saffron House, 6–10 Kirby Street, London EC1N 8TS.

Any person who does any unauthorized act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The authors have asserted their rights to be identified as the authors of this work in accordance with the Copyright, Designs and Patents Act 1988.

First published 2015 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndsmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St. Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries

ISBN: 978-1-137-34454-0

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

The Palgrave handbook of philosophical methods / [edited by] Chris Daly,
University of Manchester, UK.

pages cm

Includes bibliographical references and index.

ISBN 978-1-137-34454-0

1. Methodology. I. Daly, Chris, editor.

BD241.P3155 2015

101—dc23

2014049660

Contents

List of Figures

vii

Acknowledgments

viii

Notes on Contributors

ix

1

Introduction and Historical Overview

Chris Daly

Part I Philosophical Inquiry: Problems and Prospects

33

1 A Priori Analysis and the Methodological A Posteriori

David Braddon-Mitchell

51

2 The Failure of Analysis and the Nature of Concepts

Michael Huemer

77

3 Singular Ontology: How To

Alexis Burgess

112

4 Paradigms and Philosophical Progress

M. B. Willard

133

5 Disagreement in Philosophy

Jason Decker

158

6 Agnosticism about Ontology

Chris Daly and David Liggins

Part II Philosophical Explanation and Methodology in Metaphysics

179

7 Modality, Metaphysics, and Method

Boris Kment

208

8 Explanation and Explication

Paul Audi

231

9 Empirically Grounded Philosophical Theorizing

Otávio Bueno and Scott A. Shalkowski

258

10 Et Tu, Brute?

Sam Baron

282

11 Properties are Potatoes? An Essay on Ontological Parsimony

Nikk Effingham

List of Figures

12	Advice for Eleatics <i>Sam Cowling</i>	306
13	Pragmatism without Idealism <i>Robert Kraut and Kevin Scharp</i>	331
Part III Intuition, Psychology, and Experimental Philosophy		
14	Intuitions, Conceptual Engineering, and Conceptual Fixed Points <i>Matti Eklund</i>	363
15	Thought Experiments and Experimental Philosophy <i>Joachim Horvath</i>	386
16	Rationalizing Self-Interpretation <i>Laura Schroeter and François Schroeter</i>	419
17	Reclaiming the Armchair <i>Janet Levin</i>	448
Part IV Method, Mind, and Epistemology		
18	Placement, Grounding, and Mental Content <i>Kelly Trogon</i>	481
19	Theory Dualism and the Metalogic of Mind-Body Problems <i>T. Parent</i>	497
20	Knowing How and 'Knowing How' <i>Yuri Cath</i>	527
21	Philosophy of Science and the Curse of the Case Study <i>Adrian Currie</i>	553
22	Three Degrees of Naturalism in the Philosophy of Science <i>Paul Dicken</i>	573
Part V Metaethics and Normativity		
23	Against Pluralism in Metaethics <i>Jens Johansson and Jonas Olson</i>	593
24	Directly Plausible Principles <i>Howard Nye</i>	610
25	Moral Inquiry and Mob Psychology <i>James Lenman</i>	637
26	The Methodological Irrelevance of Reflective Equilibrium <i>Tristram McPherson</i>	652
Index		
		675

- Stich, S. & Weinberg, J.M. (2001). 'Jackson's Empirical Assumptions', *Philosophy and Phenomenological Research*, 62(3), pp. 637–643.
- Swain, S., Alexander, J. & Weinberg, J. (2008). 'The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp', *Philosophy and Phenomenological Research*, 76(1), pp. 138–155.
- Unger, P. (1968). 'An Analysis of Factual Knowledge', *The Journal of Philosophy*, 65(6), pp. 157–170.
- Vaidya, A. (2010). 'The Epistemology of Modality'. In E. N. Zalta, ed. *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*. Available at: <http://plato.stanford.edu/archives/sum2010/entries/modal-epistemology/>.
- Velleman, J.D. (2000). 'On the Aim of Belief.' In *The Possibility of Practical Reason*. Oxford: Clarendon Press, pp. 244–281.
- Wedgwood, R. (2007). *The Nature of Normativity*. Oxford: Clarendon Press.
- Weinberg, J. et al. (2012). 'Intuition & Calibration', *Essays in Philosophy*, 13(1). Available at: <http://commons.pacificu.edu/eip/vol13/iss1/15>.
- Weinberg, J., Nichols, S. & Stich, S. (2001). 'Normativity and Epistemic Intuitions', *Philosophical Topics*, 29, pp. 429–460.
- Williamson, T. (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Williamson, T. (2007). *The Philosophy of Philosophy*. Malden, Mass.: Blackwell.
- Williamson, T. (2011). 'Philosophical Expertise and the Burden of Proof', *Metaphilosophy*, 42(3), pp. 215–229.
- Yablo, S. (1993). 'Is Conceivability a Guide to Possibility?', *Philosophy and Phenomenological Research*, 53(1), pp. 1–42.

16

Rationalizing Self-Interpretation

Laura Schroeter and François Schroeter

A characteristic form of philosophical inquiry seeks to answer a 'what is it?' question. When philosophers ask such questions, they are looking for an informative analysis of the nature of the topic in question: what does it take for something to be knowledge? or a morally right action? or an instance of free will? or a member of a biological species? or the individual Barack Obama? or the logical function of negation? Different philosophical theories propose specific analyses of the nature of familiar but imperfectly understood topics. Alternatively, a theory will seek to show that, contrary to initial appearances, there is no single topic that we've been talking about with the relevant terms: contextualists about knowledge, for instance, argue that we pick out different epistemic statuses with the term 'knowledge' on different occasions of use, while incompatibilists about free will argue that our notion of freedom involves incoherent metaphysical commitments. In this chapter, we ask how philosophers do and should adjudicate debates about the correct answer to these 'what is x?' questions.

Our starting point in thinking about such questions is the first-person perspective of a rational inquirer. It's important to notice that from the first-person perspective, the object-level question 'what is x?' is equivalent to the meta-level question 'what is the reference (or, more generally, the semantic value) of my term "x"?' This meta-level formulation is important in part because some answers to the question – including contextualism, relativism, and some forms of error theory – simply cannot be formulated in object-level terms. But more importantly, we'll argue that meta-level considerations about one's past representational practice with 'x' are relevant to answering the 'what is x?' question even when one advocates a simple realist analysis. In this respect, our methods for answering a philosophical 'what is it?' questions are unlike our methods for answering more mundane questions like 'is x F?' or 'does x exist?'

Taking our best first-person epistemic practices seriously, we suggest, supports a distinctive meta-level approach to philosophical methodology. We'll argue

that philosophers do and should rely on a distinctive type of pragmatic and meta-representational reasoning – a form of rationalizing self-interpretation – in answering ‘what is x?’ questions. The bulk of this chapter is devoted to articulating and defending this rationalizing self-interpretation account of philosophical methodology (section 2–section 4). By working through a specific example, we seek to isolate the relevant inputs into deliberation and the methods for adjudicating between competing answers to a ‘what is x?’ question. We suggest that the self-interpretive methods we isolate generalize across the board as a way of identifying the real nature of familiar topics.

Before taking up this task, however, we place this methodological discussion within a broader theoretical framework (section 1). We posit a necessary connection between epistemic methodology and metasemantics. On our view, the correct semantic assignment for a representation must be *justifiable* from the epistemic perspective of the subject herself, given suitable empirical information and cognitive powers. So the upshot of your own ideal, fully informed epistemic methods will determine the correct answers to ‘what is x?’ questions.

This means our account of self-interpretation plays two important theoretical roles: it’s both a core constraint on the determination of semantic values (reference), and an epistemological theory about the best methodology for discovering defining characteristics of familiar topics. The best methods for answering ‘what is x?’ questions must get one closer to the truth about x, *on the original meaning associated with the token representation ‘x’* used in posing the question. It’s crucial that there be no change of meaning in the course of answering the object-level question. So an account of our best methodology cannot afford to ignore the question of how representations acquire and change their semantic values. An advantage of an approach like ours, which satisfies a justifiability constraint, is that it provides a clear account of how epistemology and metasemantics are integrated.

Although it is not universally accepted, a justifiability constraint on the assignment of semantic values remains the dominant approach to metasemantics. In recent years, prominent champions of this sort of justifiability constraint have argued that it grounds *a priori* conceptual analysis. We believe this argument rests on a mistaken account of our best epistemic methods for identifying the nature of a familiar topic, x. In the last section of the chapter (section 5), we explain how our method of rationalizing self-interpretation undercuts the case for a *a priori* conceptual analysis.

1 Justifiability

In this section, we motivate and articulate a generic justifiability constraint linking ideal epistemology and metasemantics. The core idea motivating a

justifiability constraint is that the correct semantic assignment for a representation must be justifiable from the epistemic perspective of the user of that representation.¹

Such a constraint has figured prominently in the work of leading metasemantical internalists:

The possibility of this sort of [two-dimensional] analysis of our concepts is grounded in the following general feature of our concepts. If a subject possesses a concept and has unimpaired rational processes, then sufficient empirical information about the actual world puts a subject in a position to identify the concept’s extension. (Chalmers and Jackson 2001, 323)

One might call this intuition a principle of speaker’s authority with respect to reference determining mechanisms. Or better: the principle that we are Masters of our Meanings. Gone are the days when we think that there is any first person authority with respect to reference, and even intension. But there is a remaining intuition that we are masters of our meanings in the final instance, in that what it takes to *determine* the reference of our words at least is up to us. (Braddon-Mitchell 2004, 149)²

Nor is it just internalists who find a justifiability constraint attractive. Many of the original externalist arguments appeal to readers’ own epistemic commitments about what they’re prepared to discover about the objects, kinds and properties picked out by their words (Putnam 1970, 1975; Kripke 1980; Burge 1979, 1986, 1996). In this spirit, Stephen Yablo, a critic of internalism, advocates what he calls a ‘psychoanalytic standard’ for attributing errors about modal content:

A cruder sort of doctor might say, here is how the illusion arises, take my word for it. But I would never dream of asking you to take my word for it. No, the test of my explanation is whether you can be brought to accept the explanation, and to accept that your [prior] judgment is to that extent unsupported. (Yablo 2006, 334)

It’s not hard to see why a justifiability constraint on the assignment of semantic contents has been so popular. After all, there is a core structural relation between epistemology and metasemantics: the semantic value of the subject’s use of ‘x’ determines which of the subject’s potential answers to the ‘what is x?’ question are true. A correct metasemantical theory, in other words, determines the normative goal for first-person deliberation about ‘what is x?’ questions. Moreover, as we noted above, our best epistemic methods for answering ‘what is x?’ questions have implications for metasemantics. It’s not

just that from the first-person perspective, object-level 'what is x?' questions are equivalent to metasemantic questions about the semantic value of 'x'. Our deliberation about such questions must respect certain metasemantic facts: e.g. we must ensure that changes in our substantive commitments about x in the course of inquiry don't shift the meaning of the target expression 'x', and we must be ready to consider the possibility that 'x' may shift its reference in different contexts of use. So our best epistemology of reference seems to entail metasemantic commitments. Similarly, metasemantical theorizing seems to entail commitments about first-person epistemology. The correct assignment of semantic values to an individual subject's use of 'x' must capture something important about the first-person epistemic perspective of the subject herself on the topic picked out by 'x'. Otherwise semantic values couldn't play their characteristic role in assessing the truth of the subject's beliefs and in rationalizing explanations of her actions. A justifiability constraint explains this convergence between the first-person epistemology of reference and the objective determination of semantic values by requiring that the two theories generate the same verdicts, given the same empirical inputs.

In effect, a justifiability constraint amounts to a kind of interpretive charity. Consider the case of an external interpreter who advocates a particular interpretation of your word 'x'. If you were informed about all the relevant empirical facts and arguments that grounded that interpretation, let's suppose, ideal epistemic reflection would lead you to different verdicts from the interpreter about which objects count as x. In such a case, surely your own idealized verdicts about what counts as x should trump those of the external interpreter: if the proposed interpretation cannot be justified from your point of view, it seems uncharitable to take it to determine the reference, and thus the normative standards of truth and correctness for your thoughts. Given the idealizations we have allowed, it seems arbitrary for the explanatory priorities of an external observer to trump your own epistemic priorities in determining these normative standards for your thoughts.³

The justifiability constraint must be spelled out carefully, however, if we want to accommodate the familiar externalist point that subjects can be partially ignorant or mistaken about the defining characteristics of familiar objects, kinds, or properties. What a justifiability constraint requires is simply that the subject be able to recognize her mistakes and correct them, if given the relevant empirical information and reasoning powers. The basic structure of the constraint can be formulated as follows:

(JUSTIFIABILITY) The subject's initial state of understanding when using 'x' puts her in a position to know *what x is* (so she could know which things, actual or possible, count as instances of x) on the basis of (i) (non-question-begging) empirical information about her circumstances and (ii) ideal epistemic methods.

A full characterization of Justifiability, however, must fill in conditions (i) and (ii) more explicitly.

Consider the full information condition (i). One important issue that arises in cashing out this constraint is which empirical facts are to be presented to the subject and how they are described. To avoid trivialization of the constraint, the empirical facts must be specified without using 'x' or obvious cognates, and without invoking semantic or metasemantic facts about 'x'. At the same time, the constraint should allow the subject to take into account whatever empirical considerations strike her as relevant in ideal deliberation.

David Chalmers (2006, 2012) has articulated and defended a specific approach to characterizing the empirical information relevant to deliberation about 'what is x?' questions. According to Chalmers, we can describe a complete possible world and the subject's notional location within that world, using an ideal vocabulary that specifies the ultimate supervenience-base properties for that world. This approach has the advantage of not pre-judging which empirical properties might prove epistemically relevant to the subject and of ensuring non-circularity for virtually any ordinary language expression. Of course, this approach requires considerable idealization in order to engage with subjects' ordinary epistemic methods: no finite human being would be capable of grasping an exhaustive base-level description of a possible world. If we had reason to suppose that the idealization in question would distort the subject's epistemic commitments, then this would undermine the rationale for accepting Justifiability, but see (Chalmers 2012, ch. 3) for defense of this idealization from this charge. Chalmers' framework also assumes that one can exhaustively characterize one's own empirical circumstances without relying on any empirical presuppositions – a controversial assumption (Stalnaker 2008, ch. 5.4). However, we'd like to emphasize that Justifiability need not be beholden to this controversial assumption: as long as it's possible to specify, on a case-by-case basis, those empirical facts that would strike the subject as relevant in ideal deliberation, Justifiability will impose a substantive constraint on the assignment of semantic contents. The constraint itself does not require that there be a universal descriptive framework for characterizing semantically relevant aspects of one's empirical circumstances for every possible representation.⁴

The second condition in Justifiability invokes a subject's ideal epistemic methods. These methods should reflect the subject's own epistemic commitments as a rational epistemic agent. We won't assume that the relevant methods can be specified naturalistically, e.g., as determined by the subject's dispositions in certain naturalistically specifiable circumstances or by some purely formal method. Whether ideal epistemic methods are naturalistically reducible in this way will be left open. Given our finite and fallible reasoning dispositions, it would be implausible to think the subject's epistemic commitments can be simply read off from her actual dispositions: like the empirical inputs

into deliberation, the subject's commitments about methodology will involve idealizing actual dispositions. We'll consider the structure of these epistemic commitments in section 3.

Before we do so, we'd like to briefly consider two objections to Justifiability. First, one might object that Justifiability amounts to a type of verificationism, since it ties semantic values to the subject's own criteria for deciding what falls into a word's extension. But verificationism seems to impose implausible psychological constraints on what our words and thoughts can represent. So Justifiability may seem to threaten robust realism about familiar objects, kinds, and properties. However, such worries are misplaced. Justifiability idealizes away from the subject's psychological limitations, with respect to both empirical information and computational capacity. Suppose there are some empirical facts that are in principle unverifiable – say, facts about unobserved objects or about quiddities. Justifiability requires us to respect the subject's reasoning about how these facts, if they obtain, affect which things count as *x*. So Justifiability is consistent with the possibility of unverifiable empirical facts. Similarly, even if the subject's actual psychological dispositions are finite and unreliable – e.g. subjects cannot comprehend facts above a certain level of complexity and many subjects are disposed to affirm the consequent – Justifiability requires us to respect the verdicts that the subject would arrive at if she were reasoning according to *ideal epistemic methods* (reflecting her own epistemic commitments). Given these two idealizations, the subject's actual psychological limitations simply drop out of the equation. What's left is a claim about how the facts about the subject's circumstances constrain the facts about the semantic content of her words and thoughts; the correct interpretation must be justifiable by ideal epistemic methods, which reflect the subject's own epistemic commitments.⁵

Second, one might worry that Justifiability suffers from the opposite problem: far from being too restrictive, the constraint may seem so permissive as to be vacuous. Until we have a precise theoretical specification of the relevant epistemic methods, it may seem that Justifiability can't rule out any semantic interpretations at all. Of course, we agree that 'ideal epistemic methods' stands in need of further clarification. But even in its current formulation, which relies on an intuitive specification of epistemic methods, Justifiability is substantive enough to rule out many semantic interpretations. The key notion of epistemic justifiability, after all, is not infinitely fungible: there are some verdicts about what *x* is that simply cannot be justified from the subject's perspective by any intuitively correct methods. There is no way, for instance, that it could turn out that what it takes to be a frog is to be a prime number; and no amount of empirical information or fancy reasoning should convince you that cows are really cosmic rays. So at the very least, Justifiability rules out such absurd, subjectively indefensible semantic assignments. And even in its current

form, Justifiability rules out some metasemantical theories. Ruth Millikan, for instance, explicitly denies that a subject's current understanding must put her in a position to identify the correct reference of her own words and thoughts after ideal reflection (Millikan 1984, 1993). Georges Rey has argued that *all* naturalistic, broadly causal theories of reference will be vulnerable to what he calls the 'fortuitous locking' objection: the designated causal referential relation will lock a representation onto an intuitively unjustifiable referential candidate (Rey 1992). Far from being vacuous, Justifiability rules out some – perhaps all – reductive causal theories of reference.

In the next sections, we focus on our epistemic methods for resolving questions of the form 'what is *x*?' Using a test case, we'll ask how one should adjudicate among competing verdicts to such a question: what considerations could justify one answer over others?

2 Inputs into deliberation

What is water? Or equivalently, what is the reference of 'water' as you currently understand that term? In order to bring to light the epistemic principles governing our deliberation about such questions, it's helpful to consider how we might adjudicate between competing answers to this question.

We'll consider two such answers that have enjoyed support in the philosophical literature. The current philosophical orthodoxy is that water is H₂O (Putnam 1975). This is intended as a straightforward identity claim: the kinds picked out by your ordinary use of the term 'water' and by your chemical term of art 'H₂O' are one and the same. This orthodox view has been challenged by theorists like Mark Johnston, who argue that water is a manifest kind, whose nature is tied to superficially observable features that are irrelevant to the identity of the underlying chemical kind (Johnston 1997). Proponents of the manifest kind view are a bit vague about just which observable features are relevant to defining the kind (e.g. its distinctive color, smell, potability; its role in constituting rain, lakes, rivers, and oceans; its suitability for cooking or washing; etc.) and about how exactly these features help to demarcate the manifest kind. Following prevailing philosophical convention, we'll use the term 'watery stuff' as a place-holder for some more specific specification of a manifest kind constitutively tied to macro-level properties that guide your everyday thinking about water.⁶ Like the orthodox view, the manifest kind view makes a property identity claim: your ordinary use of 'water' and the term of art 'watery stuff' pick out one and the same kind. The two views thus propose incompatible answers to the same 'what is it?' question.

What are the best epistemic methods for adjudicating this dispute? What sorts of considerations would epistemically justify the claim that one of these views about the nature of water is true? We distinguish two aspects of this

question. In this section, we ask about the inputs into deliberation: which commitments about water should you take into account when trying to determine what water really is? In the following section, we consider the best epistemic methods for updating your beliefs about the nature of water in the light of these inputs.

When trying to determine the essential defining characteristics of water, you start in *medias res* with a complex set of beliefs and implicit assumptions about water. To fix ideas, consider the following incomplete but representative list of typical commonsense assumptions about water among contemporary English speakers:

- *Particular instances*: there's water in this bottle, in Port Phillip Bay, Lake Michigan, etc.
- *Perceptual Gestalts*: the characteristic look, taste, odor, tactile resistance and heaviness of water.
- *Physical roles*: water's rough boiling point, its transformation into steam, its role as a solvent, the fact that it expands when it freezes, etc.
- *Biological roles*: water's necessity for the survival of plants and animals; how it's ingested; the effects of water deprivation; etc.
- *Practical roles*: the roles water plays in agriculture, transport, washing, cooking, surfing, etc.
- *Symbolic roles*: water is strongly associated with cleanliness and purity, it plays an important role in many religious rituals, etc.
- *Explanatory roles*: water has a non-obvious explanatory structure, which explains many of its characteristic roles; water is composed of H₂O.
- *Epistemology*: water is easy to spot but hard to define; our beliefs about water may be mistaken or incomplete; observation of instances of water grounds induction to unobserved cases.

Which of these commitments should you take into account in deliberation about the real nature of water? How widely should you cast your deliberative net?

We suggest that *all* these commitments are potentially relevant inputs into deliberation about what water is. It would be a mistake to restrict the inputs into deliberation to your own current attitudes, excluding your past assumptions and those of others in your community. Still less should you restrict your attention to some core subset of convictions that currently strike you as particularly obvious or central. After all, when you ask what water is you're interested in determining *what all of us have been thinking and talking about all along*. And the topic in question, you assume, is *genuinely of mutual interest*. You're not simply trying to find out what matches your own current idiosyncratic conception of water right now. So your verdict should be justifiable in the light of

the totality of opinions about water within your linguistic community, past and present – even those you yourself don't currently accept.

This expansive view of the relevant inputs into deliberation, we suggest, is grounded in and justified by basic psychological facts about our ways of keeping track of a topic in thought and talk. We've argued elsewhere that our most basic cognitive mechanisms for keeping track of particular topics in thought ('mental files') commit us to historically-extended representational traditions, in which all appropriately connected thoughts are presumed to represent the very same topic: from your first-person perspective, your current question about the nature of water *seems guaranteed* to pertain to the very same topic as the 'water' thoughts of your past self and of others in your linguistic community (Schroeter 2012; Schroeter and Schroeter 2014). So *all* thoughts in the same tradition will strike you as potentially relevant to justifying a verdict about the nature of water.

In the remainder of this section, we unpack these claims. We first explain what we mean by a subjective appearance of guaranteed sameness of topic that extends throughout a shared representational tradition, and we then explain why this subjective appearance justifies an expansive view about the inputs into deliberation.

From the first-person perspective, certain thoughts seem to pertain to the same topic *de jure*: it seems obvious and indubitable that there's just one topic in question. When you start consciously deliberating about what water is, for instance, it may occur to you *that water is potable* and *that water boils at 100°* and *that there's water in Port Phillip Bay* and *that water is good for your health*, and so on. Throughout this conscious episode of reasoning, your various thoughts seem logically guaranteed to pertain to the very same thing: it *seems* as if there's no logical possibility that the water that's potable might fail to be the water that boils at 100°. Given your way of thinking about things, the question of sameness seems closed, in a way that *de facto* identity claims like 'water is H₂O' or 'water is watery stuff' do not: even when you accept such non-trivial identities, they seem rationally dubitable. This appearance of *de jure* sameness, we suggest, is generated by your most basic cognitive mechanisms for keeping track of a single topic in thought – your mental filing system. Mental files are basic cognitive mechanisms that bind together a bundle of attitudes and cognitive dispositions in such a way as to generate this immediate appearance of *de jure* sameness of topic (Schroeter 2012).

Your mental filing system, moreover, doesn't just establish fleeting *ad hoc* relations among co-conscious thought episodes. Mental files establish stable dispositional relations among your standing attitudes. Normally, your standing belief about water's potability will not become disconnected from your belief about water's boiling point, and get linked instead to a belief about vodka's boiling point. If apparent *de jure* sameness relations were unstable in this way,

you could never build up any coherent and stable system of beliefs about a topic. Relations of apparent *de jure* sameness reflect stable and reliable links among your standing attitudes that automatically dispose you to treat linked attitudes as pertaining to the same topic.

These dispositional relations, in turn, generate the appearance of *de jure* sameness of topic among your thoughts over time. For instance, you may have a memory of learning that water can form ice crystals as a child. This memory is richly linked to your other attitudes about water right now: e.g. it seems to pertain *de jure* to the same topic as your beliefs about water's boiling point and potability. The memory itself also seems to pertain *de jure* to the very same topic as the past judgment from which it derives. Indeed, when you consciously entertain such a memory, it seems as if you're simply reactivating the very same attitude with the very same content as you entertained in the past – so the question of sameness of topic seems closed. Your memories thus extend the appearance of *de jure* sameness from your current set of 'water' attitudes into the past. Moreover, the link to your past attitudes is not just a matter of one or two long-standing memories. At each stage in your history, your current attitudes will be richly linked to your own immediate prior attitudes by such relations, forming a continuous chain of apparent *de jure* sameness.

The interpersonal case is structurally similar. If you're a normal English speaker, you'll automatically hear others' use of 'water' as expressing attitudes pertaining to the very same topic that you yourself associate with that term. And as long as you rely on your automatic mechanisms for linguistic understanding, the question of sameness of topic seems closed: from your point of view, it immediately seems obvious and logically guaranteed that there's just one topic in question whenever you discuss 'water' with others – there's no possible way the world could be in which water isn't water. We suggest that the explanation of this appearance is grounded in your mental filing system: just as your own standing attitudes are stably bound together by your mental files, others' mental files are stably linked to yours in virtue of individuals stably associating their own mental files with the word 'water' in the public language (cf. Cumming 2013). The direct link via your mental filing system explains why your thoughts strike you as standing in direct logical relations to others'. Moreover, your commitments to direct logical relations among thought contents aren't just limited to your own immediate interlocutors. Your current attitudes are linked by apparent *de jure* sameness relations to those of your interlocutors, and their attitudes in turn are linked to those of others in the community. So your own thoughts are richly linked via chains of apparent *de jure* sameness to a network of coordinated thoughts, past and present, of others within your linguistic community.

Starting from your current thoughts about water, we can trace back these diachronic and social chains of apparent *de jure* sameness to demarcate a

continuous, historically extended *representational tradition*. From your epistemic perspective, there *seems* to be a direct logical guarantee of identity of topic at each stage in this representational tradition. Within your own thoughts, this immediate appearance of stable *de jure* sameness makes sense of your readiness to take recently unearthed memories at face value, as pertaining *de jure* to the same topic as your current 'water' thoughts. And within your community, this presumption of stable *de jure* sameness makes sense of your readiness to hear new interlocutors using the word 'water' as making claims that pertain *de jure* to the same topic as your own 'water' thoughts.⁸ In this way, the basic cognitive structure of your thoughts – the cognitive mechanisms that allow you to keep track of a topic – gives rise to *prima facie* epistemic appearances of *de jure* sameness that naturally dispose you to assume *de jure* sameness of topic throughout the representational traditions to which you belong.⁹

Given this default presumption of *de jure* sameness of topic throughout your representational tradition, what follows about the relevant inputs into ideal epistemic deliberation about 'what is x?' questions? When you deliberate about the nature of water, you'll take your answer to apply not just to your own 'water' thoughts but to the 'water' thoughts of everyone in your representational tradition. Your aim is to figure out what all of us have been thinking and talking about all along. Given this aim, it would be arbitrary to confine the inputs to your deliberation to your own current assumptions about water, ignoring your own earlier views and those of others in your community. Why take your own current opinions to settle the question about what's been represented by everyone in the shared representational tradition? Surely you ought to take into account what others have thought about the topic, given that your verdict is intended to apply to everyone. After all, you are just one finite and fallible thinker among many: others may have access to important insights about water that you have missed, and your own understanding of what's important may be idiosyncratic, limited, or misguided. Given the default assumption that we're all coordinating on the same topic, therefore, it would be epistemically irrational to exclude others' attitudes as potentially relevant inputs into deliberation about what water is. This is not to say that others' opinions are necessarily correct – given our conflicting opinions, it's clear that at least some of us must be mistaken. Nor is it to say that the majority opinion must be correct – in many cases, it makes sense to conclude that the majority is mistaken about the nature of what's represented by our shared representational traditions. It's just that insofar as your verdict purports to characterize what all of us have been thinking about all along, you're committed to that verdict being warranted by the totality of our shared representational practice.

In sum, our suggestion is that it's rational to treat the representational tradition associated with a 'what is x?' question as the default *unit of interpretation*.

This tradition demarcates the set of attitudes that are relevant inputs into your deliberation about the nature of x; and in deliberation you seek to characterize the content of all of these attitudes. So the representational tradition demarcates, at least provisionally, both the inputs and outputs of deliberation about 'what is x?' questions.

3 Deliberative methods

Let's turn now to the ideal epistemic methods for answering such questions. How should you adjudicate between competing verdicts about the nature of water? What type of reasoning would support the conclusion that water is a manifest kind rather than a chemical kind (or vice versa)?

Notice that the correct answer to a 'what is x?' question plays an important normative epistemic role: it determines the ultimate standards for evaluating 'water' beliefs in your representational tradition as true – and hence as acceptable. So in asking what water really is, you're asking about the ultimate standards of acceptability to which you should hold your beliefs accountable. Your aim is to identify the mutually relevant topic that settles the ultimate standards by which you evaluate your 'water' beliefs as true and acceptable. From the first-person perspective of the deliberator, moreover, you don't seem to have any neat algorithm for identifying the nature of water. Before you explicitly reflect on the question of what water is, your own assumptions about the topic are bound to be heterogeneous, incomplete, and partially contradictory – and this heterogeneity is only exacerbated when you take your whole community's views into account. Thus justifying an answer to a 'what is x?' question is nothing like slotting some missing values into an implicitly grasped formula. Your goal in rational deliberation is to find some principled way of prioritizing and systematizing your own and your community's commitments about water, so as to identify the appropriate normative standards for evaluating the truth and acceptability of beliefs about the topic.

Our proposal is that ideal epistemic methods for answering 'what is x?' questions hinge on rationalizing interpretation of one's representational traditions. You need to diagnose the most important representational interests at stake in a representational tradition with 'x', and you should identify the correct verdict about the nature of x as the one that makes best sense of those interests.

Consider how you might justify the claim that water is really a manifest kind, watery stuff. Proponents of this view privilege the *everyday practical interests* subserved by classifying stuff as water: e.g. we use water for drinking, cooking, cleaning, agriculture, transport, swimming, and so on. Arguably, these practical interests are much more important to sustaining and justifying our shared representational tradition of thinking about water than any interest we may have in identifying water's underlying chemical structure. After all,

the overwhelming majority of our beliefs, desires and intentions about water concern the manifest macro-level properties that make water suited to fulfilling our everyday practical projects. In contrast, commitments about the chemical composition of water plays a relatively minor role in our representational tradition: only a small group of scientists have any real interest in analyzing the chemical properties of water, and their scientific projects are a recent development within our well-established commonsense tradition of thinking about water. If the dominant interests at stake in classifying stuff as water are everyday practical ones, rather than systematic scientific ones, then it's plausible that water is a manifest kind constitutively tied to macro-level observable properties. In that case, anything that looks, tastes, cleans and nourishes like water belongs to the very same kind that we've been picking out with the term 'water': so XYZ on Putnam's Twin Earth would count as genuine water.

This principle – that the kind picked out by our term 'water' must fulfill our everyday practical interests – is not-specific enough to demarcate the boundary between water and non-water. To get a more precise specification of the intension of 'water', we'd need to know more about the relative importance of our different practical interests and empirical information about which interests are mutually satisfiable. But even in its present form, the principle gives us reason to reject the view that water is identical to the micro-level chemical kind H_2O , since this kind seems ill-suited to vindicating our everyday practical interests in the category. A single molecule of H_2O , for instance, cannot play any of the practical roles that ground our interest in water (cf. Johnston 1997: 579). Moreover, H_2O molecules formed from the rare isotopes deuterium (2H) or tritium (3H) are unsuited to our ordinary interests: e.g. both heavy water (2H_2O) and radioactive super-heavy water (3H_2O) are poisonous to plants and animals if taken in sufficient quantities. Nor can we identify water with quantities of pure 1H_2O , for that would entail there is hardly any unadulterated liquid water and the category would be useless for everyday practical purposes (van Brakel 1986; Needham 2000, 2011; Weisberg 2005; Leslie 2013). These familiar arguments against the orthodox view are supported by the assessment of the dominant practical interests at stake in our representational tradition.

How might a proponent of the orthodox view respond? Defending the orthodoxy, we suggest, hinges on showing that underlying *explanatory interests* are actually more central to justifying and sustaining our shared representational tradition than everyday practical interests. You might plausibly argue that our everyday practices of using water in cooking, cleaning and agriculture, as well as the incremental technological advances we make in these practical areas, all presuppose that water has a stable underlying explanatory structure. A burgeoning literature in developmental and experimental psychology on

'psychological essentialism' lends support to this view: children as young as two seem to tacitly assume that all instances of a natural kind like water share intrinsic, non-manifest properties that causally explain its manifest observable properties. This essentialist assumption explains a wide range of inductive, explanatory and classificatory dispositions of both children and adults.¹⁰ So an interest in water's causal explanatory role is not confined to a small sub-community of scientists: all of us have a strong background interest in the explanatory role of water. Modern chemistry simply extends and systematizes these commonsense explanatory interests, allowing for greater prediction and control at the familiar macroscopic level. A proponent of the orthodox chemical interpretation of 'water' could argue that this explanatory interest constitutes the real, unifying justification for our shared representational practice with 'water'.

Of course, it might turn out that a dominant commonsense interest in a unifying hidden essence is not vindicated by chemistry.¹¹ If chemical kinds are always defined purely at the micro-level, then there is no single chemical kind that plays the essentialist roles we associate with water. In that case, it may make sense to revise your verdict about what's most important to our representational tradition. If there is no single plausible filler of the core explanatory roles we take water to play, then perhaps our explanatory interests are less central to sustaining and justifying our representational tradition than they seemed before we knew the results of scientific inquiry. Arguably, our explanatory commitments about water aren't dominant enough to justify the conclusion that there's no determinate kind that we've been thinking and talking about all along. Alternatively, one might argue that there really is a unified macro-level chemical kind that vindicates our explanatory interests: to be an instance of water just is to be a macro-level system composed predominantly of H_2O molecules. Of course there will be chemically important variations between such systems at the micro-level, depending on whether the system is in a liquid, solid, or gas phase, and depending on which other chemical components are part of the system, etc. But a proponent of the orthodox view can plausibly argue that these details do not detract from the main moral that our presumption that water has a unified core explanatory structure has been vindicated.¹²

Whether it's correct to accept the orthodox view thus depends in part on the relative importance of explanatory interests vs. purely practical interests in our representational tradition of classifying stuff as water. And its correctness also depends on controversial empirical issues in the philosophy of chemistry about the individuation of chemical kinds. Our primary concern here, however, is not whether the orthodox interpretation of 'water' is in fact right, but rather which type of reasoning would support this interpretation over other alternatives.

It's worth briefly considering a third interpretation: perhaps 'water' is ambiguous between a chemical kind and a manifest kind, picking out different kinds on different occasions of use (Lewis 1994). It's important to see that an ambiguity interpretation doesn't win by default: to support this view, you must establish that an ambiguous interpretation is more plausible than the competing univocal interpretations. Moreover, your argument will take the same general form as those we've already considered. First, you must show that the two types of representational interests at stake – practical and explanatory – are equally important. Second, you must show that these interests are not in fact mutually satisfiable: there's no single semantic interpretation that vindicates them both. Lastly, you must show that an ambiguity interpretation is not ruled out by our initial presumption of *de jure* sameness. As we noted in section 2, the presumption of *de jure* sameness of topic that there's just one topic in question is central to induction, inquiry, and debate about water. According to the ambiguity interpretation, these appearances are misleading: we've been thinking about different topics on different occasions and we've been implicitly treating these distinct kinds as if they were one. Whether this is an acceptable interpretation, we suggest, depends on how important the presumption of stability of topic has been to justifying our representational tradition. Does the fact that we've been systematically conflating two distinct topics undercut any justification for our past categorizing practices? If so, the ambiguity interpretation should be abandoned in favor of an error theory. But if we still see our past practices as locally justified ways of keeping track of two importantly distinct topics, perhaps with some isolated incidents of conflation, then it's reasonable to interpret that past practice as ambiguous.¹³

We think the lessons about our epistemic methods for answering 'what is water?' sketched in section 3 generalize to other cases. For a discussion of further test cases, see the appendix.

4 First-person epistemology of reference

Let's take stock. Our example suggests a *pragmatic meta-cognitive epistemology* for answering 'what is x?' questions. Ultimately, supporting a specific verdict about the nature of a familiar topic involves (i) identifying and prioritizing the practical and theoretical interests subserved by your representational tradition, and (ii) refining your substantive understanding of the topic so as to best meet those interests. With semantic ascent, this method amounts to an ideal first-person epistemology of reference: it tells us how to determine the defining characteristics of the basic topics (objects, kinds, properties, etc.) we keep track of in thought and talk. Thus as rational epistemic agents we seem committed to a pragmatic meta-cognitive epistemology of reference.

This analysis rests on three core claims about the reasoning in our examples. First, on our account the ideal epistemology of reference is *meta-cognitive*. Rationally adjudicating between radically different verdicts about the nature of water (race, free will, color, content) requires you to focus on meta-level facts about your own and others' cognitive states. As we argued in section 2, a representational tradition as a whole constitutes the default unit of interpretation: you're not just interested in what satisfies your own current conception of water – you want to know what you and your community have been thinking and talking about all along. Since the representational tradition linked with your use of 'water' includes attitudes and practices that you don't currently endorse and situations you aren't currently in, an ideally justified verdict will be based in part on your meta-level beliefs about the attitudes and practices within that tradition, including those of others and of your past self. So the *inputs* into ideal deliberation about the nature of water are not just your current substantive conception of water together with ideal object-level information about your physical environment: you must take into account meta-level facts about the whole representational tradition associated with your current 'what is x?' question.¹⁴ Moreover, your *methods* in ideal reasoning are also essentially meta-cognitive. You cannot simply trust your current substantive conception of water to determine what counts as water, given empirical facts about your social and physical environment.¹⁵ If the starting point for deliberation is the default assumption that the unit of interpretation is a stable representational tradition, then you should admit that your current substantive conception of water may fail to reflect important aspects of your representational tradition. To determine the ultimate standards that you (and others in your representational tradition) should hold your beliefs accountable to, you need to take a step back from your current substantive beliefs and consider how to interpret the point of the representational tradition as a whole. Thus ideal methods for adjudicating 'what is x?' questions are essentially meta-cognitive and interpretive: you're not just relying on your current conception of water to reason about the world, you're stepping back from your current conception and reasoning about how to interpret a shared representational tradition.

Second, ideal methods for answering 'what is x?' questions involve *pragmatic* forms of reasoning. When interpreting your representational tradition, you construe it as a practice with a *point* or *rationale*: a set of categorizing and inferential dispositions that allow you and other participants in your linguistic community to keep track of a topic that's of mutual interest. More specifically, you evaluate which representational interests are most important to justifying and sustaining your representational practice with 'x' and you engage in means/end reasoning about how those interests are best satisfied. Both aspects of this type of reasoning are characteristic of pragmatic reasoning about what to do. In reasoning about 'what is x?' questions, however, you focus on the

distinctively representational interests at stake in your representational traditions and on the ultimate epistemic standards for evaluating beliefs.

Which representational interests are most important, and thus relevant to identifying the nature of the topic, is settled through *rationalizing interpretation* of the representational tradition. Relative importance isn't determined by the consciously accessible beliefs or intentions of participants in the tradition. Nor is it determined by straightforward causal explanations of how the tradition became entrenched. From the perspective of a rational epistemic agent, these straightforward ways of identifying interests would arbitrarily cut off deliberation about the nature of the things you're thinking about. The question of whether water is a manifest or an explanatory kind, for instance, can't be conclusively settled by pointing to empirical facts. Even if there was a widespread prior belief about the relative importance of practical vs. theoretical interests to our thinking about water, this belief could turn out to be mistaken: we aren't infallible about what's most important to us. Similarly, what's most important causally isn't always what's most important epistemologically. If our everyday practical interests turned out to be more important to causally entrenching our representational tradition of thinking about water, one could still argue that our explanatory interests are more important to settling what exactly we're thinking and talking about.¹⁶ From the deliberative perspective of a rational epistemic agent, the interests that are relevant to adjudicating 'what is x?' questions are those that help justify or *rationalize* that tradition. Ideal methods for adjudicating 'what is x?' questions don't simply construe representational practices as meeting psychologically or causally fixed representational interests. Our interpretive methods construe them as meeting representational interests that help make sense of our practices – that help construe them as having a point or rationale.¹⁷

The third point we want to emphasize is that this interpretive method is genuinely *epistemological* – it helps us get closer to the truth about the nature of the topics we're thinking and talking about. Many theorists will want to challenge this claim. Given that the interpretive method involves pragmatic and rationalizing considerations, one might worry that it tells us about what we *should* be thinking and talking about, not what we *are* thinking and talking about. This worry, however, is misplaced.

Recall that our starting point in this chapter was Justifiability: the answer to 'what is x?' must be justifiable using the subject's own ideal epistemic methods. Given this constraint, an objector must establish that we're wrong about the ideal epistemic methods for adjudicating 'what is x?' questions and thus determining the truth and acceptability conditions of your beliefs about x. We've used a specific example to illustrate how ideal reflection appeals to pragmatic meta-cognitive reasoning about the point of our representational traditions (see the appendix for some further examples). Given the plausibility of these

methods from the first-person perspective, the objector owes us an explanation of why their apparent epistemological authority is illusory. What exactly are the allegedly better epistemological methods for discovering the real nature of the objects, kinds and properties you think about? And why should you take those alternative methods to determine the truth and acceptability of your beliefs?¹⁸

In fact, there is a strong reason to reject the skeptic's appeal to the is/ought distinction – for it fails to take into account the *epistemic ambition* inherent in our first-person perspective. As rational epistemic agents, we normally take our words and thoughts to represent *genuinely interesting and important features* of the world – not just whatever happens to satisfy our current criteria. When asking about the nature of water (or free will, color, etc.), we don't assume that we (or our community as a whole) already implicitly know the right answer. We initially launch a representational practice with some rough and ready criteria for identifying a seemingly important but imperfectly understood topic (water, color, free will, etc.). In revising our initial understanding of that topic, moreover, we don't rely on some determinate, implicitly grasped guidelines for inquiry. Instead, rational inquiry into the nature of the topic seeks to vindicate the initial open-ended presumption that we're picking out a topic that's genuinely of mutual interest. And what's genuinely interesting in the context of systematic philosophical or scientific inquiry, we suggest, depends on facts about the most important representational interests subserved by our past representational tradition. What's most important in fixing the reference of 'water' shouldn't be understood in a purely perspective-independent way: for our words and thoughts don't always pick out perspective-independent "joints of nature" that are important "from the point of view of the universe" – nor should we want them to. For the purposes of determining the semantic contents of thought and talk, what's interesting and important must be understood as relative to the subject's own projects and concerns associated with a topic. The method we have sketched precisely aims at determining what's interesting and important relative to the subject's own past representational tradition. So from the point of view of a rational epistemic agent, these pragmatic meta-cognitive methods are ideally suited to getting us closer to the truth about the interesting and important topics that we were thinking and talking about all along.

5 Justifiability without *a priori* conceptual analysis

In closing, we'd like to briefly explain why, if we're right about ideal epistemic methods, Justifiability doesn't support *a priori* conceptual analysis.

In a series of influential publications, David Chalmers and Frank Jackson have argued that a justifiability constraint on the determination of semantic values grounds *a priori* conceptual analysis (Chalmers and Jackson 2001;

Jackson 1998; Chalmers 1996, 2006, 2012). The idea is that if a subject has the ability to identify the nature of the topic picked out by her terms on the basis of full empirical information and ideal reflection (as Justifiability requires), then she'll have *a priori* access to material conditionals of the form 'if Twin Earth is actual, then water is XYZ', which capture the application conditions of her term 'water' relative to any conceivable way the world could be. The reason access to these conditionals is *a priori*, they claim, is that all the relevant empirical information can be packed into the antecedent of the conditionals. So knowledge of the conditionals themselves does not depend for its justification on any knowledge of facts about the subject's actual empirical circumstances.

If we are right about ideal methods for answering 'what is x?' questions, then this case for *a priori* conceptual analysis fails. Chalmers and Jackson assume that deliberation about 'what is x?' questions involves exclusively object-level reasoning: you simply use your substantive understanding of x to come to verdicts about what counts as x in any possible situation. But the meta-cognitive aspect of rationalizing self-interpretation entails that your ideal verdicts about 'what is x?' questions depend essentially for their justification on *a posteriori* knowledge of facts about your actual belief about x and the representational tradition to which they are linked. In a nutshell, the problem is that you cannot discharge all of the empirical presuppositions you rely on in self-interpretive reasoning into the antecedent of the conditional. You may of course apply the interpretive methods we've outlined to a purely hypothetical representational tradition to generate a verdict about what participants in that tradition are talking about with their use of 'x'. But if you want to justify your own object-level verdicts about what x is – verdicts that use your own word 'x' to express your commitments about the topic x, rather than merely *mentioning* 'x' in characterizing the commitments of some hypothetical subject – then you're committed to the claim that your actual 'x' representation that's used in deliberation exists and is appropriately causally and descriptively related to the overall representational tradition that you're interpreting. So your meta-cognitive epistemic methods depend essentially for their justification on empirical knowledge about the existence and properties of the representational states you're actually using in deliberation.¹⁹

In response, one might argue that there's a simpler route to vindicating *a priori* conceptual analysis. Perhaps it's essential to competence with the concept you currently express with 'water' that you be disposed to accept certain application conditionals after ideal reflection. These application conditionals would then count as *analytic* or *conceptual truths*: any change in your actual understanding that would alter your ideal verdicts about application conditionals would *eo ipso* count as a change in concept. Moreover, the *a priori* status of these conditionals would be secured by the account of conceptual competence. So even if your justification for accepting a judgment of the form 'if Twin Earth is actual, then water is XYZ' does rest on meta-cognitive empirical knowledge

of your *actual* 'water' thoughts, this judgment is still guaranteed to be true on pain of changing the concept.

This defense of a *a priori* conceptual analysis, however, rests on a dubious claim about what's required for competence with the same concept. In order to deploy the same concept over time (or between subjects), one's ideal verdicts about 'what is x?' questions must remain stable relative to every conceivable scenario considered as actual. We've argued elsewhere that this criterion for sameness of concept does not reflect the rational epistemic agent's own commitments about *de jure* sameness of topic over time or between subjects.²⁰ Alternative models of concept identity based on causal-historical relations (such as representational traditions) explain *de jure* sameness of topic without positing a precise match in ideal verdicts about application conditionals – and they do so in a way that better reflects our best epistemology. On such relational accounts, Justifiability doesn't entail analytic or conceptual truths that are *a priori* in virtue of conceptual competence conditions.

To sum up. The self-interpretive methods we've defended here reflect the open-ended character of the first-person epistemology of reference. Justifiability requires semantic assignments to be justifiable from the subject's ideal perspective, but it does not in itself say anything about the methods of justification or the inputs into deliberation. It's tempting to simply assume that this justification must be grounded in some initial pattern of substantive understanding of the topic, x, that functions as epistemic foundations for answering 'what is x?' questions. That is, the individual's current understanding prioritizes among her different object-level assumptions about x in such a way as to provide a *priori* constraints on the determination of semantic values for her 'x' thoughts, no matter what the actual world turns out to be like. These internal constraints would ground a *priori* conceptual analysis – for the very same constraints that guide ideal inquiry would also constrain the assignment of semantic values. But our account of ideal epistemic methods for answering 'what is x?' questions shows that Justifiability is not wedded to this individualistic and foundationalist picture of the epistemology of reference. In fact, such a picture fails to vindicate the epistemic ambition characteristic of the first-person epistemology of reference. In ideal deliberation, you should take the reference to depend, not on your own current substantive conception of the topic, but on the conception that best meets the most important representational interests that justify and sustain the associated representational tradition. This commitment to your actual representational tradition, we have suggested, introduces ineliminable meta-cognitive presuppositions that preclude a *priori* conceptual analysis. And your epistemic commitment to categories that are of genuine mutual interest undermines the individualist foundationalist picture of justification. Your own current substantive understanding of x – including the way you're currently disposed to prioritize among your different substantive commitments – may

be overturned by elements of the tradition about which you are ignorant or mistaken. Thus, taking Justifiability seriously in fact grounds an anti-foundationalist and anti-individualist account of reference determination.

Appendix: further examples

Consider the philosophical debate about the nature of race. What are we attributing when we characterize people as black or white? In answering this question, some theorists treat *biological explanatory interests* as most important: both children and adults tend to treat racial categories as base-level biological kinds, positing internal properties passed from one generation to the next that ground a wide range of inductions and explanations of manifest properties. Since biological theory does not vindicate the relevant sorts of essentialist explanations, such theorists often argue that racial categories are empty.²¹ In contrast, social constructivists about race focus on the role that racial categories play in social domination: classifying someone as 'black' situates that person in a social hierarchy and activates a suite of negative stereotypes and discriminatory norms. This interest in *locating individuals within a social hierarchy* grounds very different interpretations of the nature of racial categories: being black or white does not depend on biological characteristics – instead, it is a marker of social status tied to certain manifest properties.²²

Next consider the nature of free will and moral responsibility. One might argue that the dominant interest at stake in our representational practice with these categories hinges on a certain model of *causal explanation*: what's most important to our practice is distinguishing actions whose ultimate causal source is the agent herself from actions that can be explained without mentioning the agent. If this is the point of our representational practices with 'free' and 'responsible', then we'll need to choose between libertarianism or an error theory about the properties picked out by these terms (depending on the actual facts about causation). Alternatively, one might argue that the dominant interest at stake in categorizing actions as free and agents as responsible is in *regulating social interactions and fostering pro-social behavior*, in part via reactive attitudes such as praise and blame. If so, then some form of compatibilism seems plausible: the precise extension of our terms will depend on how the relevant interests are best fulfilled, given the empirical facts about our psychology and social institutions.

We can discern a similar structure in the case of color. One might argue that our dominant interest in keeping track of colors is to capture a *rich theoretical role derived from visual appearances*: a color such as red seems to be a uniform surface property shared by all red objects and light sources, which causes red experiences, and which stands in certain relations of similarity and difference to other colors such as blue, green, and yellow (e.g. as represented in a color

wheel). If this theoretical role really is crucial to justifying and sustaining our interest in categorizing objects as red or blue, then we'll be pushed towards an error theory of color since nothing actually satisfies this role. But one might argue that what's most crucial to justifying and sustaining our categorizing practice with 'red' is a *more limited causal explanatory role*: we're interested in keeping track of an underlying physical property of objects (integrated with the rest of physical theory) that causes red sensations in normal circumstances. If this is our dominant interest, then the actual physical facts might justify a disjunctive physicalist account of what it is to be red. Alternatively, one might argue that our dominant interest in color categories hinges more on *visual salience* than on hidden physical structure: the primary reason we're interested in categorizing things as red is that red things all share a similar immediately accessible visual appearance. If perceptual salience is our dominant interest in color categories, then arguably a (rigidified) dispositional account of color would be justified.

Lastly, consider the philosophical debate over the nature of representational content and intentionality – a debate that bears directly on our concerns in this chapter. Everyone agrees that the content of a person's words and thoughts plays an important role both in explaining her behavior and in setting standards for assessing her attitudes for truth and rationality. But these explanatory and normative interests tend to pull theories in different directions, and theorists tend to privilege one over the other. Naturalistic theories of content privilege the *causal explanatory role* of content attributions as most important: although information theoretic and teleosemantic theories identify very different types of explanatory roles for content, they agree that our primary interest in content ascriptions is their role in the systematic scientific explanation of action. Traditional descriptivist theories and interpretivist theories, in contrast, privilege *norms of rationality*: what exactly your words and thoughts represent depends on what it would be rational for you to do and which sentences it would be rational for you to accept and reject. Proponents of these theories tend to hold that the causal explanatory role of representational contents is *rationalizing* explanation, which cannot be assimilated to scientific explanatory paradigms based on causal correlation or natural selection. There are, of course, deep differences between these rationality-based approaches. In particular, descriptivists privilege the subject's current substantive understanding: they seek to isolate some core descriptive criteria that allegedly guide a subject's inquiry in answering 'what is x?' questions come what may. In contrast, interpretivists privilege the point of view of an external observer who seeks an interpretation of the totality of the subject's behavioral and linguistic dispositions that construes her as rational overall. Still, these very different types of theory implicitly agree about the dominant rationale or point of intentional content ascriptions: they both take ideal norms of rationality, rather

than causal generalizations, as decisive in determining the content of a subject's words and thoughts.

Our own approach to semantic content in this chapter falls in this broad camp of those who privilege norms of rationality over causal explanation. We take the central interest in representational content to be that it settles normative standards for the truth and acceptability of belief. But our account cannot be fully assimilated to either the descriptivist or traditional interpretivist positions. We agree with the descriptivist that the relevant normative standards should respect the subject's own ideal epistemic perspective – Justifiability captures this assessment – but we disagree with the descriptivist's assumption that the individual grasps some topic-specific reference-fixing criterion that determines how she should justify answers to 'what is x?' questions. Our view about the ideal methods guiding first-person deliberation is similar in some respects to those advocated by third-person interpretivists. But unlike standard interpretivism, our account of interpretive methods is motivated by norms for epistemic self-governance rather than by an external rationalizing explanatory perspective. In sections 2 and 3 we have sought to remind readers through the use of examples of the inputs and methods that rational epistemic agents take to be relevant to resolving questions about the nature of familiar topics. We discuss the basis for our disagreement with the descriptivist tradition, and internalism more generally, in section 5.

Notes

1. Our focus here will be on words whose primary semantic role is clearly representational – words that pick out empirical objects, kinds, or properties. We'll call the semantic content of such words *the reference*. On this usage, the reference is not simply the expression's actual extension – it is the intuitive subject matter picked out by the word that determines its *modal profile* or *intension*. Of course, not all words have a reference in this sense. Some expressions whose primary semantic role is representational fail to secure reference to anything: in such cases, we'll say the word has an empty referential content. Many expressions play non-representational semantic roles: e.g. they may contribute to quantification and binding, or they may play various inferential or expressive roles, or they may express relations among propositions, or they may serve as conventional place-holders for a contextually-determined reference. We believe that the epistemological methods we advocate can be extended to determine the semantic contents of non-representational expressions, but for present purposes we will confine our attention to representational content.
2. Justifiability constraints have been advocated by a host of different theorists working in the traditional, broadly Fregean approach to reference determination (e.g. Frege, Russell, Dummett, Evans, Peacocke, Carnap, Jackson, Chalmers). Gareth Evans, for instance, accepts 'Russell's Principle' for determining reference to individuals: in order to think about a particular object, one must know which object is in question, where this is understood as a discriminatory ability to uniquely identify the object in question (Evans 1982, 89). In a similar spirit, Christopher Peacocke suggests that

the correct assignment of a semantic value to any concept must make the belief-forming practices mentioned in its possession conditions correct – i.e. true or truth-preserving (Peacocke 1992, 20 ff.). Recently, David Chalmers has articulated and defended different versions of the 'Scrutability of Truth', which requires the ideal justifiability of any arbitrary sentence, not just those that purport to answer 'what is x?' questions (Chalmers 2004, 2006, 2012).

3. Of course, not all philosophers agree that normative facts must be subjectively justifiable in this way. In particular, some metaethicists hold that normative facts may be rationally inaccessible to the subject (Enoch 2011; McDowell 1998, ch. 5). On this view, your epistemic starting point could prevent you from ever recognizing the correct normative standards of reference and truth-conditions governing your words and thoughts: you simply cannot get there from here – at least not by relying on your own best epistemic methods and base-level empirical information. But normative theorists who reject a justifiability constraint face an important challenge: they must explain why subjectively unjustifiable normative standards are genuinely authoritative for the subject herself. It would be a significant theoretical cost to leave it mysterious why some standards are authoritative for individual subjects. A justifiability constraint, in contrast, underwrites a straightforward explanation of why normative standards bind a subject: they are the ones the subject would accept if ideally informed and ideally reflective.

4. If there is no universal base-level vocabulary whose representational content can be entertained independently of any empirical presuppositions, then justifiability (together with a specification of the base vocabulary) won't provide a complete theory of intentionality. Even so, justifiability imposes a substantive and intuitively correct constraint on the semantic values of the subject's own representational system. Justifiability can constrain an externalist theory of intentionality in two different ways. On a bottom-up approach, the externalist can appeal to an independent externalist account of the content of some class of base-level representations (e.g. a teleosemantic theory of perceptual contents) and then use justifiability to constrain the assignment of semantic contents to more complex representations. On a holistic approach, the externalist can appeal to general principles of rationalizing interpretation to assign semantic contents to the representational system as a whole, given the system's history and circumstances.

5. David Lewis made a similar point about appealing to an ideal interpreter in his account of radical interpretation:

It should be obvious by now that my problem of radical interpretation is not any real-life task of finding out about Karl's beliefs, desires, and meanings. I am not really asking how we could determine these facts. Rather: how do the facts determine these facts? By what constraints, and to what extent, does the totality of physical facts about Karl determine what he believes, desires, and means? (Lewis 1974, 333–4)

6. We needn't assume that there is a finite definition of the relevant kind couched in terms of the subject's ordinary language vocabulary. The kind might, for instance, be specified in part by invoking subjects' classificatory dispositions or similarity space. And the definition of a manifest kind may have vague boundaries, where it is indeterminate whether a case counts as an instance of the kind. Still the manifest kind view is committed to the claim that there is a fact of the matter about which things, actual and counterfactual, count as instances of the kind (and which things are penumbral cases). See (Chalmers 1996, ch. 2.4; Jackson 1998, ch. 2–3) for discussion of these points. As we'll see, defending the claim that water is a manifest kind, rather than a scientific kind does not hinge on having a precise definition available.

7. Frege highlighted this phenomenon when he contrasted trivial and non-trivial identity claims involving proper names (Frege 1892). We introduced the related contrast between *de jure* and *de facto* sameness of topic to highlight two facts: (i) what's at issue is merely a subjective appearance, and (ii) this appearance directly concerns sameness or difference of topic, not substantive understanding of that topic (Schroeter 2008, 2012; Schroeter and Schroeter 2014). It's worth noting that the appearance of *de jure* sameness is not always veridical (Schroeter 2007). You can of course raise a substantive question of sameness of topic by resorting to *meta-level descriptions* of your object-level train of thought. For instance, you might ask whether *all those thoughts* really refer to the same thing. Or you might resort to scare quotes, asking whether the 'water' in Port Phillip Bay is the same stuff as the 'water' that boils at 100°. What you cannot do is simply redeploy the very same object-level perspective on water involved in your earlier train of thought if you want to raise a non-trivial identity question.

8. There are, of course, significant differences between the intra- and inter-personal cases. Unlike basic cognitive filing mechanisms involved in mental files, linguistic competence must be acquired: the appearances of *de jure* sameness among speakers will only be generated among fluent speakers. Relatedly, the mechanisms for ensuring semantic coordination between individuals are less secure than within an individual. As a consequence, we are more prepared to step back from an object-level appearance of *de jure* sameness with others' thoughts than with our own past or present thoughts. Still, this is a difference in degree and salience, not a difference in kind: one cannot rule out the possibility of equivocation even in one's own thoughts (Schroeter 2012).

9. Why assume that transitivity of apparent *de jure* sameness holds within the representational tradition? We suggest this is a natural consequence of the fact that the apparent *de jure* sameness is the appearance of logically guaranteed identity of topic. Since identity is transitive, the appearance of identity commits one to transitivity. However, it's important to keep in mind that this is a mere appearance: the default assumption that there is a direct logical guarantee of sameness can be defeated by empirical evidence of a shift in topic.

10. For an overview and defense of psychological essentialism see (Gelman 2003, 2004); see (Strevens 2000) for a critique.

11. (Leslie 2013) argues that we are naive scientific essentialists about kinds like water, but that this assumption is in fact false.

12. See (Hendry 2006) for defense and refinement of the orthodox view along these lines.

13. An *ambiguity interpretation* entails that particular past thought episodes were unambiguously about one of the two kinds. (This is, after all, the way that ordinary ambiguity works.) For this disambiguating approach, see (Schroeter 2007, 2008), for similar positions see (Kitcher 1978; Burge 1998). An alternative way of responding to the situation of having two equally important but mutually unsatisfiable types of representational interests, is to advocate an *indeterminacy interpretation*, according to which each thought episode expressed by 'water' has indeterminate reference: in Field's terminology it 'partially refers' to watery stuff and it 'partially refers' to H₂O (Field 1972). So how are we to evaluate the truth of claims with indeterminate reference? Field opts for supervenient semantics. Others might argue that the lack of determinate reference supports an error theory.

14. One might object that the meta-representational beliefs needn't figure essentially in your reasoning methods. Perhaps, after identifying the relevant attitudes, you could simply incorporate them into your own object-level perspective in a kind

of 'make believe' or 'simulation' of the communal perspective, and then engage in straightforward object-level reasoning about the nature of water from that simulated perspective. We don't see any support for this methodology in actual reasoning about disputed cases. Moreover, the suggestion faces daunting internal problems. Any extended representational tradition will involve disagreements, but you cannot simply add logically conflicting attitudes to your own object-level commitments about water. If your purely object-level reasoning about the nature of water is based on contradictions, then its justificatory status will be severely compromised. The rational response to discovering contradictory commitments within one's own perspective is to step back from actively endorsing the contradictory propositions, and instead to take up a metacognitive perspective in deliberating about how to reconcile one's contradictory commitments in the light of one's total evidence.

15. Here we take ourselves to disagree with the position articulated by (Chalmers 2012); see (Schroeter forthcoming) for discussion.
16. Consider a parallel case. Perhaps an interest in social domination was more important to *causally explaining* how our representational tradition with 'witch' became entrenched than a purely theoretical interest in magical explanation. From the first-person epistemic perspective of deliberating about what witches are, our theoretical interest in explaining malevolent magical powers may nonetheless be more important to determining what it is to be a witch.
17. It's worth emphasizing that rationalizing interpretation will not necessarily vindicate a representational practice. For instance, the most important interests subserved by a representational tradition won't always be satisfied – as the witch case illustrates. And in some cases, the most important interests subserved may be interests that we shouldn't have – for instance, because they are morally objectionable, as illustrated by the interest in social domination in the race case.
18. Theorists like Frank Jackson and David Chalmers draw a sharp distinction between the semantic value 'x' *actually does have* and the semantic value 'x' *should have* given the interests at stake in your past representational tradition. Like us, these theorists accept justifiability as a constraint on the assignment of reference for ordinary thought at talk. Moreover, these theorists also agree that, *as philosophers*, we should be concerned with answering the normative question: what's really interesting and important is which objects kinds or properties best meet our representational interests (Jackson 1998, 44–46; Chalmers 2011, 538). However, these theorists deny that such philosophical methods are relevant to identifying what's picked out in ordinary thought and talk. We take this distinction between philosophical and commonsense epistemic methods to be ill-founded. As philosophers, we seek to elucidate what's of genuine interest within our shared representational traditions. So the best epistemic standards governing philosophical inquiry *just are* the best epistemic standards governing ordinary conceptual practices.
19. For further elaboration of this line of argument, see (Schroeter 2005, 2006, forthcoming).
20. See (Schroeter 2003, 2008, 2012; Schroeter and Schroeter 2014) for an elaboration and defense of a relational account of conceptual competence based on representational traditions. For an alternative relational account of sameness of concept and meanings based on a naturalistic reproduction relation and natural selection, see (Millikan 1984). Relational approaches to sameness of meaning are not new, but they have been gaining prominence in recent years. Prominent advocates of relationally individuated meanings or concepts include (Perry 1980, 2001; Burge 1993,

1998; Kaplan 1990; Cumming 2013), while (Fine 2007; Pimillos 2011) defend an irreducible semantic relation of *de jure* sameness that does not vindicate a strict meaning-identity relation.

21. (Appiah 1996; Zack 2002) argue for eliminativism about standard racial concepts for these sorts of reason. In contrast, (Kitcher 2007) argues for a modest biological account of race that defines racial categories in terms of broad genealogical lineages that ground superficial biological kinds (analogous to different breeds in dogs).
22. See (Haslanger 2000; Mallon 2006) for two versions of social constructivist theories of race.

References

- Appiah, Anthony. 1996. 'Race, Culture, Identity: Misunderstood Connections', in *Color Consciousness*, edited by A. Appiah and A. Gutman. Princeton NJ: Princeton University Press.
- Braddon-Mitchell, David. 2004. 'Masters of Our Meanings', *Philosophical Studies* 118:133–152.
- Burge, Tyler. 1979. 'Individualism and the Mental', *Midwest Studies in Philosophy* 5:73–121.
- . 1986. 'Intellectual Norms and Foundations of Mind', *Journal of Philosophy* 83:697–720.
- . 1993. 'Content Preservation', *Philosophical Review* 102:457–488.
- . 1996. 'Our Entitlement to Self-Knowledge', *Proceedings of the Aristotelian Society* 96:91–116.
- . 1998. 'Memory and Self-Knowledge', in *Externalism and Self-Knowledge*, edited by P. Ludlow and N. Martin. Palo Alto, CA: CSLI Publications.
- Chalmers, David. 1996. *The Conscious Mind*. New York: Oxford University Press.
- . 2004. 'Epistemic Two-Dimensional Semantics', *Philosophical Studies* 118:153–226.
- . 2006. 'The Foundations of Two-Dimensional Semantics', in *Two-Dimensional Semantics: Foundations and Applications*, edited by M. Garcia-Carpintero and J. Macia. Oxford: Oxford University Press.
- . 2011. 'Verbal Disputes', *Philosophical Review* 120:515–566.
- . 2012. *Constructing the World*. Oxford: Oxford University Press.
- Chalmers, David, and Frank Jackson. 2001. 'Conceptual Analysis and Reductive Explanation', *Philosophical Review* 110:315–61.
- Cumming, Samuel. 2013. 'From Coordination to Content', *Philosophers' Imprint* 13 (4):1–17.
- Enoch, David. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.
- Evans, Gareth. 1982. *The Varieties of Reference*. New York: Oxford University Press.
- Field, Hartry. 1972. 'Theory Change and the Indeterminacy of Reference', *Journal of Philosophy* 70:462–81.
- Fine, Kit. 2007. *Semantic Relationalism*. Oxford: Blackwell.
- Frege, Gottlob. 1892. 'On Sinn and Bedeutung', in *The Frege Reader*, edited by M. Beaney. Oxford: Blackwell (1997).
- Gelman, Susan A. 2003. *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford: Oxford University Press.
- . 2004. 'Psychological Essentialism in Children', *Trends in Cognitive Science* 8:404–409.

- Haslanger, Sally. 2000. 'Gender and Race: (What) Are They? (What) Do We Want Them To Be?', *Noûs* 34:31–55.
- Hendry, Robin Findlay. 2006. 'Elements, Compounds, and Other Chemical Kinds', *Philosophy of Science* 73:864–875.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
- Johnston, Mark. 1997. 'Manifest Kinds', *Journal of Philosophy* 94:564–583.
- Kaplan, David. 1990. 'Words', *Proceedings of the Aristotelian Society* supp. vol. 64:93–119.
- Kitcher, Philip. 1978. 'Theories, Theorists, and Theoretical Change', *Philosophical Review* 87:519–547.
- . 2007. 'Does "Race" Have a Future?', *Philosophy and Public Affairs* 35:293–317.
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Leslie, Sarah-Jane. 2013. 'Essence and Natural Kinds: When Science Meets Preschooler Intuition', in *Oxford Studies in Epistemology*, edited by T. Gendler and J. Hawthorn. Oxford: Oxford University Press.
- Lewis, David. 1974. 'Radical Interpretation', *Synthese* 23:331–344.
- . 1994. 'Reduction in Mind', in *A Comparison to the Philosophy of Mind*, edited by S. Guttenplan. Oxford: Blackwell.
- Mallon, Ron. 2006. 'Race: Normative, Not Metaphysical or Semantic', *Ethics* 116:525–551.
- McDowell, John. 1998. *Mind, Value, and Reality*. Cambridge, MA: Harvard University Press.
- Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
- . 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.
- Needham, Paul. 2000. 'What Is Water?', *Analysis* 60: 13–21.
- . 2011. 'Microessentialism: What Is the Argument?', *Noûs* 45:1–21.
- Peacocke, Christopher. 1992. *A Study of Concepts*. Cambridge, MA: MIT Press.
- Perry, John. 1980. 'A Problem About Continued Belief', *Pacific Philosophical Quarterly* 61:317–332.
- . 2001. *Reference and Reflexivity*. Palo Alto, CA: CSLI Publications.
- Pimillos, N. Angel. 2011. 'Coreference and Meaning', *Philosophical Studies* 154:301–324.
- Putnam, Hilary. 1970. 'Is Semantics Possible?', in *Languages, Belief and Metaphysics*, edited by H. E. Kiefer and M. K. Munitz. New York: SUNY Press.
- . 1975. 'The Meaning of "Meaning"', *Minnesota Studies in the Philosophy of Science* 7:131–193.
- Rey, Georges. 1992. 'Semantic Externalism and Conceptual Competence', *Proceedings of the Aristotelian Society* 92:315–333.
- Schroeter, Laura. 2003. 'Gruesome Diagonals', *Philosophers' Imprint* 3 (3):1–23.
- . 2005. 'Considering Empty Worlds as Actual', *Australasian Journal of Philosophy* 83:331–347.
- . 2006. 'Against A Priori Reductions', *Philosophical Quarterly* 56:562–586.
- . 2007. 'The Illusion of Transparency', *Australasian Journal of Philosophy* 85:597–618.
- . 2008. 'Why Be an Anti-Individualist?', *Philosophy and Phenomenological Research* 77:105–141.
- . 2012. 'Bootstrapping Our Way to Samesaying', *Synthese* 189:177–197.
- . forthcoming. 'Scrutability and Epistemic Updating: Comments on Chalmers' *Constructing the World, Analysis*.
- Schroeter, Laura, and François Schroeter. 2014. 'Normative Concepts: A Connectedness Model', *Philosophers' Imprint* 14 (25):1–26.
- Stalnaker, Robert. 2008. *Our Knowledge of the Internal World*. Oxford: Clarendon Press.
- Strevens, Michael. 2000. 'The Essentialist Aspect of Naive Theories', *Cognition* 74:149–175.
- van Brakel. 1986. 'The Chemistry of Substances and the Philosophy of Mass Terms', *Synthese* 69:291–324.
- Weisberg, Michael. 2005. 'Water Is Not H₂O', in *Philosophy of Chemistry: Synthesis of a New Discipline*, edited by D. Baird, E. Scerri and L. McIntyre. New York: Springer.
- Yablo, Stephen. 2006. 'No Fool's Gold: Notes on Illusions of Possibility', in *Two-Dimensional Semantics*, edited by M. Garcia-Carpintero and J. Macia. Oxford: Oxford University Press.
- Zack, Naomi. 2002. *Philosophy of Science and Race*. New York: Routledge.