

# Action-First Attitudes

Jack Marley-Payne

MMATHPHIL Mathematics and Philosophy University of Oxford, 2010

Submitted to the Department of Linguistics and Philosophy  
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY  
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
September 2016



© Jack Marley-Payne. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

<sup>n.m.</sup>  
**Signature redacted**

Signature of Author: \_\_\_\_\_ Department of Linguistics and Philosophy

<sup>h' r' i' s</sup>  
**Signature redacted**

Certified by: \_\_\_\_\_

<sup>1</sup>  
**Signature redacted** \_\_\_\_\_  
Robert Stalnaker  
Professor of Philosophy  
Thesis Supervisor

Accepted by: \_\_\_\_\_  
Roger White  
Professor of Philosophy  
Chief, Committee on Graduate Students



# Action-First Attitudes

Jack Marley-Payne

Submitted to the Department of Linguistics and Philosophy  
On September 5, 2016 in Partial Fulfillment of the Requirements  
For the Degree of Doctor of Philosophy in Philosophy.

## ABSTRACT

In this thesis, I present an *action-first* theory of knowledge and belief. We have a mutual interest in the successful action of our peers, and the significance of belief and knowledge stems from their role in promoting this success. Knowledge states tend to guide successful action, in an appropriately systematic manner. Belief states systematically guide our *attempts* to achieve our goals, and would lead to success if all went well.

In defending the action-first account, I draw on a kind of *pragmatism*: we should look to the practical role of belief and knowledge attribution, in a social setting, to determine the nature of belief and knowledge themselves. The action-account states that the role of knowledge attribution is to identify and promote successful agents. This implies that knowledge itself is a state that tends to guide successful action.

Similarly, the role of belief attribution is to help us predict how people will attempt to achieve their goals, and correct them to avoid failure where necessary. This implies that beliefs are action-guiding states that may not be success conducive – these are states that are apt to *become* knowledge given the appropriate evidence or argument. A final point is that the role of our ascriptions of rationality (and irrationality) is to *promote* practices that tend to lead to knowledge. This gives us a unified account of our concepts of knowledge, belief and rationality, founded in a cooperative society's interest in mutual success.

Granting the action-account leads to significant consequences in epistemology and philosophy of mind. It gives us reason to reject various accessibility principles, and accept intellectualism with regard to know-how. All states that lead to successful action in a systematic manner, even if we do not consciously endorse their content, fit with the rationale of the action-account. Further, the account suggests a new way to model conflicted mental states, and suggests rethinking the role of the Bayesian ideal in our conception of rationality.

These consequences, in turn, provide motivation for the action-account itself on pragmatic grounds: it opens up promising new lines of inquiry in philosophy.

Thesis Supervisor: Robert Stalnaker  
Title: Professor of Philosophy

## Acknowledgements

Many people have helped me write this thesis, including some, I'm sorry to say, I've very likely forgotten.

First, special thanks to my committee: Bob Stalnaker, Alex Byrne, and Agustin Rayo – who have looked through countless drafts over the 4+ years since I started producing thoughts on these topics, and helped me make countless more improvements along the way.

Second, thanks to the many people who have read through chapters, sat through presentations, and talked over my ideas with me. A woefully partial list: Bernhard Salow; Brendan de Kenessey; Dan Greco; Declan Smithies; Dylan Bianchi; Eric Schwitzgebel; Jack Spencer; Justin Khoo; Katya Botchkina; Lyndal Grant; Mark Richard; Melissa Schumacher; Michael Hannon; Paulina Sliwa; Rebecca Millsop; Ryan Doody; Sally Haslanger; Sinan Dogramaci; Sophie Horowitz; Steve Yablo.

I want to single out my team of 11<sup>th</sup> hour proofreaders: Arden Ali; Doug Kremm; Jay Hodges; Johnny Beale; Matt Mandelkern; Milo Phillips-Brown; Olivia Bailey; Rose Lenehan; and Zeynep Soysal. (And to be clear they've all also helped me with much more than errant commas at other times.)

Finally, a different kind of thanks is required for Jill, Milo, and Arden for living with me while I finished this thing off.

# Contents

<i>Introduction</i>	7
Chapter 1: Finding the Propositional Attitudes	11
Chapter 2: The Practical Role of Knowledge Attribution	39
Chapter 3: Against Elitist Theories of Belief	65
Chapter 4: Task-Indexed Belief	85
Chapter 5: Epistemic Ideals	115
Appendix: Know-How and Gettier: The Plot Thickens	145
<i>Bibliography</i>	157

[blank by design]

# Introduction

In this thesis, I present an *action-first* theory of knowledge and belief. We have a mutual interest in the successful action of our peers, and the significance of belief and knowledge stems from their role in promoting this success. Knowledge states tend to guide successful action, in an appropriately systematic manner. Belief states systematically guide our *attempts* to achieve our goals, and *would* lead to success if all went well.

In defending the action-first account, I draw on a kind of *pragmatism*: we should look to the practical role of belief and knowledge attribution to determine the nature of belief and knowledge themselves. In chapter 1, I motivate and elaborate this pragmatist methodology. I start with the following challenge: empirical psychology presents a picture of our minds involving a vast range of representational states, and we need to locate belief and knowledge within this more detailed picture. Further, in doing so we must explain the *significance* of these states, at least if they are worthy of prolonged philosophical attention.

Traditional approaches, conceptual analysis and naturalism, do not offer appealing answers. The reason for this is that they ignore the fact that knowledge and belief attribution is incredibly *useful*. By focusing on this, pragmatism offers a promising alternative. In order to justify the claim that the practical role of attitude attribution can tell us about the attitudes themselves, I end chapter 1 by sketching a theory of reference that links the *meaning* of a term to its practical role.

The practical role of knowledge ascription is commonly held to involve identifying good informants, but this approach is too restrictive: in virtue of being in a cooperative society where the success of our peers is mutually beneficial, we have an interest in identifying successful agents in general, and provision of accurate testimony is only a small part of this. I argue in chapter 2 that the practical role of knowledge ascription is to identify and promote successful agents. This focus on action leads to a relatively broad conception of knowledge. In slogan form: all robustly true sufficiently sophisticated action-guiding states are knowledge.

The practical role of belief attribution is closely related to that of knowledge. As well as identifying successful agents, we have an interest in identifying cases in which agents are apt to fail, in order to plan for their failure, or correct them if possible. This leads to a view on which beliefs are action-guiding states that may not be success conducive – they are states that are apt to *become* knowledge given the appropriate evidence or argument. The ability for belief states to be influenced provides a link with rationality. The role of our ascriptions of rationality is to *promote* practices that tend to lead to knowledge. We criticize and praise the doxastic states of our peers through such ascriptions in order to improve their epistemic practices. This gives us a unified account of our concepts of knowledge, belief and rationality, founded in a cooperative society's interest in mutual success.

Granting the action-first account leads to significant consequences in epistemology and philosophy of mind, which I explore in the later chapters of the dissertation. In chapter 3, I look at the sense in which knowledge and belief must be *sophisticated* mental states, and argue that it must be liberal enough to apply to skilled action, thus motivating intellectualism with regard to know-how. Further, it gives us reason to reject various accessibility principles, since we are not committed to subjects being able to articulate what they know. These results are the opposite to those that follow from the testifier account of the practical role of knowledge attribution.

This broad conception of belief and knowledge, leads to many cases in which subjects possess a pair of beliefs with contradictory contents. In chapter 4, I look at how we should think about these cases, granting the pragmatist methodology. I argue that pragmatic considerations require us to view belief as a three-place relation indexed to the tasks such states guide, in order for our account of belief to capture their action-guiding role in a systematic manner. This task-indexed account offers a new way of thinking about certain Frege cases, as an alternative to traditional theories of modes of presentation. In particular, it suggests a new response to the knowledge argument – one that avoids the problems that come with 'phenomenal modes of presentation.'

Finally, in chapter 5, I look at what the pragmatist conception of rationality can tell us about Bayesian ideals in epistemology. I argue that we should view the Bayesian ideal as one epistemic ideal among many, rather than the complete story, and sketch an

alternative ‘resource management ideal’ to use alongside it. This new picture allows us to resist Chalmers’ argument for the a priori, and preserve theoretically attractive level-bridging principles for higher-order evidence in the face of intuitive evidence against them.

These consequences, in turn, provide motivation for the action-account on pragmatic grounds: it opens up promising new lines of inquiry in philosophy.

[blank by design]

## Chapter 1

# Finding the Propositional Attitudes

### 1. Introduction

My goal is to argue for an *action-first* account of the propositional attitudes, on which what distinguishes the attitudes from other kinds of psychological states is that they guide action in a systematic and sophisticated manner. I will be doing so on pragmatic grounds: states with this action-guiding role are incredibly useful to know about, since they allow us to predict and influence behaviour systematically and efficiently. The first step, therefore, is to justify the idea that such pragmatic concerns are relevant to finding the correct account of the attitudes. In this chapter, I will show why traditional approaches do not appear well-placed to provide fruitful answers to the questions of concern, suggesting that it's worth looking for an alternative strategy: I claim that pragmatism looks to be a promising approach.

My starting point is the following question: where do we locate the propositional attitudes within the complex array of findings presented by empirical psychology? Our best scientific theory of the mind presents a picture of our cognitive structure that involves a great range of representational states. Various branches of psychology identify everything from conscious linguistic thoughts, to mental images, to representational states involved in pre-perceptual processing, as part of our total theory of the mind. It's not clear where the propositional attitudes fit within this picture: it's natural to think they are some subset of the total representational states found in psychology, but it's not obvious which ones. The propositional attitudes are generally held to be of particular *philosophical* significance – certainly judging by the amount they have been discussed – and one would hope that the locating them within a broader psychological framework would illuminate *why* they are important.

Philosophers have approached this issue in a number of different ways: one way is to look at where to draw the line between beliefs and ‘sub-doxastic states’; another is to look at where to draw the line between propositional knowledge and ‘non-intellectual states’. A form of this second question that has received much recent discussion is the question of whether know-how is a form of know-that. I will, for the most part, confine myself to this question in this chapter, to keep the discussion focused: the competing methodological approaches of concern are discussed more recently and thoroughly with respect to this topic, making it a good case study. It should be kept in mind, though, that my conclusions about knowledge are intended to carry over to belief – I’ll elaborate on this point in section 6 of Chapter 2.

I will first look at two familiar strategies for assessing intellectualism: *conceptual analysis*, which says we should look to our intuitions about knowledge; and *naturalism*, which says we should look for appropriate psychological kinds. I’ll argue that neither approach leads to a promising account – conceptual analysis struggles to provide a determinate answer, while naturalism is in danger of leading to eliminativism. This motivates looking for an alternative: I will show how pragmatism provides just that.

## **2. Introducing Intellectualism**

Intellectualism is the thesis that know-how is a kind of know-that. It was first examined explicitly by Ryle (1949), as a way into investigating *intelligence*. Ryle was concerned with the view – orthodoxy at the time – that all intelligent behaviour must result from conscious deliberation. He took it to be obvious that know-that involved conscious deliberation, and that know-how was intelligent. Therefore, showing that know-how was not know-that would be a step towards showing that there could be intelligence without conscious deliberation. I think this question about the scope of intelligence, though vague, is highly suggestive (and closely related to my broader question about where to locate the attitudes within our overall cognitive architecture) – it’s what makes the seemingly technical question of whether intellectualism holds of wider interest.

Before we get to that, though, we must be more precise about the target thesis (letting S be an agent,  $\phi$  an action, and p a proposition):

*Intellectualism*: For S to know how to  $\phi$  *just is* for S to know that p, for an appropriate proposition p.<sup>1</sup> (p will contain information on how one should  $\phi$ .)

This thesis is highly contentious, and it brings out a familiar tension between the role of knowledge in linguistic and conscious judgement on the one hand, and action-guidance on the other. In paradigm cases of know-that, the state fits with *both* conditions: when a subject knows that p, she will typically assert that p if asked, and this information will guide her action appropriately. For example, I know that Morpeth is a town in the north of England, and will say so if anyone asks; further, if I need to meet someone in Morpeth, I will travel to the north of England.

Typical cases of know-how provide action-guidance, but not linguistic or conscious judgment. For example, though I know how to ride a bike, I am unable to articulate many of the essential features of this activity. Until recently I would confidently assert that one must lean right when approaching a turn to the right. However, I recently learned (from Elga & Rayo (ms)) that one must initially lean *left* – and of course I had been leaning this way all along.<sup>2</sup>

The link between know-that and assertion suggests intellectualism is false, while the link between know-that and action guidance suggests it is true. Which feature takes priority will play an important role in determining the correct verdict on the matter.

To put things more precisely, the following principle is incompatible with intellectualism:

---

<sup>1</sup> The phrase ‘just is’ stands for some sort of equivalence between the two propositions flanking it. The minimal reading of this is as of having the same truth conditions (i.e. as ‘if and only if’). This will do for our purposes since even the minimal version of intellectualism is highly contested.

<sup>2</sup> It is sometimes suggested that one *can* articulate the content of one’s know-how by means of *demonstrative reference* – see, e.g., Stanley and Williamson (2001). For example, if I know-how to ride a bike, I could start riding it and say ‘*this* is how you ride a bike’. I do not think this is the appropriate kind of articulation. The state that guides one’s action does not directly cause one to make the assertion, unlike when one recalls information in response to a question – one has to go through the roundabout process of using the know-how state to act, and then referring demonstratively to the action.

*The Articulation Principle:* If one knows that p then *ceteris paribus* one will be able to assert that p when asked.<sup>3</sup>

An alternative thesis, similar in spirit, would be a ‘principle of introspection’, which states that if a subject knows that p, she must have *conscious access* to the information – that is, if one knows that p then one must be disposed to consciously endorse p.<sup>4</sup> This too seems incompatible with intellectualism, since we lack conscious access to how we carry out various skilled actions.

However, if we focus on the action-guiding aspect of knowledge we arrive at a principle that seems more hospitable to intellectualism:

*The Action Principle:* If one possesses an information state with content p which guides one’s actions appropriately, *ceteris paribus* one knows that p.<sup>5</sup>

It is plausible to think that cases of know-how involve information states. Know-how is *learned* and so involves sensitivity to environmental states of affairs. Moreover, such states guide subjects in their attempts to realize their ends. Thus, they meet the fairly weak criteria for an information state outlined by Dretske (1981) and Stalnaker (1984). That is, the state causally covaries with a particular environmental state of affairs, and guides action in a manner determined by its content.<sup>6</sup>

---

<sup>3</sup> The ‘*ceteris paribus*’ caveat is designed to allow for paralysis and related circumstances where assertion isn’t possible. I take it that unarticulated know-how cannot be classified as falling under the *ceteris paribus* clause, it seems a normal part of human psychology, rather than the product of strange external circumstances.

<sup>4</sup> This kind of principle is defended by Smithies (2012).

<sup>5</sup> The ‘*ceteris paribus*’ here rules out exotic cases of action guidance without knowledge. For example, when a mad scientist implants a behaviour-overriding probe into a subject’s brain. Note, also that this principle is closely related to the ‘knowledge-action principle’ defended by Hawthorne and Stanley (2008) – that says, roughly, that your knowledge constitutes your *reasons* for action. I’m working with a weaker principle that sets aside normative issues since they won’t be relevant in what follows.

<sup>6</sup> In cases where a subject S desires to  $\phi$  and knows how to  $\phi$  they satisfy the following conditional: if it were the case that  $w$  was a way for S to  $\phi$ , S would  $\phi$ . This means S roughly satisfies Stalnakerian criteria for knowing that  $w$  is a way for S to  $\phi$ . For more on this see Glick (2011).

Establishing either the articulation thesis or the action thesis will go a long way towards settling the question of intellectualism.<sup>7</sup> Choosing between them is tricky, though; and, as with many contentious questions in philosophy, there is a dispute over what factors determine the answer. There are two prominent approaches: on the one hand, we might look to features concerning ordinary language and armchair intuition; on the other we might look to empirical findings in psychology.

This reflects a more general controversy over the role of intuitions and that of theoretical significance in determining the nature of our concepts. The two sides of this debate are often termed *conceptual analysis* and *naturalism* respectively – terminology I’ll adopt in what follows, though I’ll usually employ the term ‘psychologism’ to refer to naturalist approaches in philosophy of mind in particular.

An important note before moving on is that I will be sketching extreme versions of each of the methodological views. It’s likely that many of the philosophers who are typically labelled as ‘conceptual analysts’ or ‘naturalists’ do not adopt such positions wholesale.<sup>8</sup> One point is that many theorists want some kind of *hybrid* approach that views intuition and theoretical significance as *both* playing a role in determining our concepts.<sup>9</sup> I will explain at the end of section 4 why I don’t think mixing and matching the two methodologies improves matters.

A second point is that both conceptual analysts and naturalists implicitly accept elements of pragmatism in their theorizing. I think, though, there is value in making this pragmatic component explicit so that the dialectic is clear and arguments are assessed in a consistent manner – where commitment to pragmatism is implicit, its principles may be applied erratically. In section 5, I discuss the significant changes pragmatism imposes on how one must pursue philosophy, so one cannot invoke it in passing, and then continue with business as usual. Moreover, a clear-eyed adoption of pragmatism leads us in new and promising directions.

---

<sup>7</sup> Granted, establishing the action principle would not establish intellectualism completely. We would still have the somewhat technical issue of moving from the claim that action guided by know-how is guided by know-that, to the claim that a know-how *state* is to be identified with a know-that state.

<sup>8</sup> Stanley and Williamson are not the remorseless adherents to conceptual analysis they are often made out to be. It’s clear from other work – e.g. Stanley (2011) and Williamson (2000) – that they take a range of non-linguistic factors to play a role in determining the nature of knowledge.

<sup>9</sup> A prominent version of this is Lewis’ (1984) eligibility theory of reference.

A final point is that I am endorsing pragmatism only with respect to the problem of locating the propositional attitudes. I am *not* claiming that it is the appropriate method for philosophy generally. Indeed, I don't think there is a single right philosophical methodology across all issues – one has to start investigating the first-order issues, and work out what research strategy is appropriate as one goes.<sup>10</sup>

### 3. Conceptual Analysis

Conceptual analysis is the methodological approach which prioritizes intuition and ordinary language when deciding whether intellectualism holds. The rough idea is that we must get clear on what we mean by the terms 'know-how' and 'know-that' in order to see if the one entails the other. This has long been the default approach in epistemology – particularly the *analysis of knowledge* in the aftermath of Gettier cases. It is natural to apply the methodology to the question of intellectualism as well. Here is a rough statement of the view:

*Conceptual Analysis:* Find the accounts of know-how and know-that which best fit with our intuitions about the concepts and see whether they validate intellectualism or not.

Determining what it is for an account to 'best fit with our intuitions' is a crucial issue. At a first pass it will have to involve making as many intuitions come out true as possible (perhaps weighting them relative to the strength/centrality of the intuition).

Stanley and Williamson (2001) present a twist on traditional versions of conceptual analysis. They investigate the *semantic structure* of know-how ascriptions, using findings from linguistics to argue for intellectualism. Though this is different in many ways from

---

<sup>10</sup> Fodor says the following about the scientific method: "I've gotten to know a number of scientists... and it's my impression that, like the rest of us, they mostly make up their methodology as they go along. The scientific method as I have come to understand it: *Try not to say anything false; try to keep your wits about you.*" [Fodor (2008) p. 4 note 3] I'm inclined to agree, and I think the same goes for philosophy. It seems to me that philosophers are often led to absurd conclusions when they become too attached to their account of what makes a good philosophical theory. For example, Lewis took himself at his word as to how one decides what objects to allow into one's ontology, which is what led him to his notorious thesis of modal realism.

traditional forms of conceptual analysis, it still focuses on ordinary language use – only it is more concerned with systematizing intuitions about what is grammatical than with first-order intuitions about cases. In any event, the objections I will consider apply to this strategy too. I will use the label ‘S+W’ to refer to dialectical bogeyman who sees this linguistic argument as decisive (rather than one factor among many).

### 3.1 Problems with Conceptual Analysis

General complaints about conceptual analysis are not hard to come by.<sup>11</sup> There are two in particular I want to consider, though – because they are especially relevant to intellectualism. The first is the worry of irrelevance; the second is the worry of indeterminacy.

The worry of irrelevance is that if conceptual analysis is correct, then it seems that whether intellectualism holds is a totally uninteresting question. On this approach, it boils down to the question of whether the best systematization of certain features of ordinary language provides an interpretation of the expressions ‘know-how’ and ‘know-that’ such that the former entails the latter. One can be hard-pressed to see how this has any significance when it comes to the nature of the mind or of rationality – and so it seems like the question is of interest only for philosophy of language, not philosophy of mind or epistemology.

In this vein, Devitt writes: “Why should we care about knowledge-how? From a naturalistic perspective, we should care only if knowledge-how is a feature of mental reality that is pertinent to psychological explanations.”<sup>12</sup> And he does not think conceptual analysis puts us onto something that meets these conditions for significance. Similarly, Kornblith says: “the subject matter of epistemology is knowledge itself, not our concept of knowledge... I do not believe that our intuitions, or our inclinations to say various things, should carry a great deal of weight in philosophical matters.”<sup>13</sup> Conceptual analysis seems to require systematizing our uninformed everyday language use for its own sake, and one might hope philosophy has something more significant to

---

<sup>11</sup> See, e.g., Laurence & Margolis (2003), Kornblith (2002), and Williamson (2008).

<sup>12</sup> Devitt (2011) p 205.

<sup>13</sup> Kornblith (2002) pp 1-2.

offer. It is highly implausible to think that this was what Ryle was *getting at* when he first asked whether know-how is a kind of know-that: it sheds no light on the broader issue of the nature of intelligence

The worry of indeterminacy is that conceptual analysis will not be able to provide stable verdicts, since virtually no two philosophers agree on what the correct intuitions concerning knowledge are. Though in some cases, such as Gettier's (1963) original examples, there is consensus as to the correct verdict; in others opinion is starkly divided. A prominent example of this is Bonjour's (1998) case of Norman the Clairvoyant, who regularly has reliably true propositions just 'pop into his head' and he believes them without being able to provide any reason as to why he thinks they are true. Philosophers are split as to whether Norman's clairvoyant beliefs constitute knowledge – those of an internalist persuasion say no, while those of a reliabilist persuasion say yes. Further examples, such as Block's (1997) 'Super Blind Sight' and Smithies' (2014) 'Hyper Blind Sight' reinforce the internalist/reliabilist divide.

The traditional concern has been deciding what *justification* condition is necessary for belief. The issue most pertinent to intellectualism, however, is whether the *articulation principle* takes priority over the *action principle* or vice-versa – as has already been mentioned. If we can't settle this matter then we have no chance of making progress on the question of intellectualism. However, I think that philosophers are divided between those who have articulation-first intuitions and those who have action-first intuitions – and hopes of a resolution are dim.

Indeed, divisions run along roughly the same lines as with the internalism/reliabilism debate about justification. This is because internalists claim (and reliabilists deny) that we must have some sort of *access* to the conditions that make a belief count as knowledge, and thus they endorse an introspection condition on knowledge – which closely related to the articulation principle.<sup>14</sup> Intuitions about specific cases are also messy and conflicted – I investigate this further in the Appendix.

One might hold out hope that further investigation will provide a solution. Perhaps there is a yet to be discovered argument that refutes one of the positions, while relying

---

<sup>14</sup> See, e.g., Snowdon (2003) for some action first intuitions, and Conee & Feldman (2001) for some articulation/introspection first intuitions. To complicate matters further, Carr (1979) claims it's intuitive to think know-how requires conscious access.

only on premises to which both sides are committed. I'm sceptical of this, however. It seems like both action-first and articulation-first theories of knowledge are internally coherent and cannot be refuted while relying only on premises their supporters find intuitively plausible.<sup>15</sup>

One might think that S+W's approach provides a way out here. Issues in linguistics are plausibly more cut and dried than first-order intuitions about knowledge. However, there are problems of a different sort. For the sake of argument, I'll grant that the linguistic results S+W cite are correct.<sup>16</sup> They show, I will assume, that the compositional semantic structure of a know-how ascription is equivalent to a certain form of know-that ascription. More precisely (1) has the semantic structure articulated by (2):

- 1) S knows how to  $\phi$ .
- 2) For some way of  $\phi$ 'ing  $w$ , S knows that  $w$  is a way for her to  $\phi$ .<sup>17</sup>

S+W wish to infer, that the fact two sentences have the same compositional structure means that they *must* express the same propositional content. This would entail that a subject possesses know-how if and only if she possesses the relevant know-that.

S+W's assumption – that the element of meaning determined by compositional semantic structure is a full-blooded proposition – is highly controversial, as has been noted by Santorio (2016). Though it is perhaps the orthodox view, it has been comprehensively challenged by so-called 'contextualists', such as Recanati (2004). They argue that the semantic structure of a sentence only determines a 'propositional radical', which lacks even truth conditions – these are provided in context by a process called 'pragmatic enrichment'.<sup>18</sup> Similarly, *expressivist* semantics for ethical language – e.g. Gibbard's (1990) account – provide a framework in which normative claims do not have truth-conditional content, despite having the same compositional structure as certain non-normative claims.

---

<sup>15</sup> Moreover, both views are at least *prima facie* reasonable – so unlike with the sceptic we cannot simply ignore them if they cannot be answered on their own terms. Cf. Williamson (2000).

<sup>16</sup> Though see Rumfitt (2003) for an argument to the contrary.

<sup>17</sup> Strictly speaking, S+W leave it open that there may also be a semantic component referring to a 'practical mode of presentation'.

<sup>18</sup> It should be noted that Stanley (2000) is a staunch critic of such views. However, I take it that his views about knowledge are meant to be independent of these criticisms.

Santorio (2016) provides a positive argument against S+W. He shows how we can accept the linguistic results S+W rely on, while giving a non-intlectualist account of know-how. He provides a semantics for knowledge ascriptions on which know-how ascriptions do not attribute the same kind of propositional knowledge as regular know-that attributions. If one accepts that it's *possible* for there to be difference in propositional content despite compositional uniformity, Santorio shows that linguistic evidence alone is not enough to vindicate intellectualism. Instead, it's natural to think we must return to conceptual analysis to decide between the semantic accounts. This is analogous to how, if one concedes that expressivist normative semantics cannot be refuted on purely linguistic grounds, one must return to ethical intuitions to argue for realism.

A second point is that even if one grants S+W's linguistic assumption, their semantic argument only tells us that a subject possesses know-how if and only if she possesses the relevant know-that – it has nothing to say about when she possesses both and when she possesses neither. We have seen that cases of skilled action which can't be articulated are intuitively described as cases of know-how but not know-that. The semantic argument implies that one of these intuitions must be wrong, but it doesn't say which. Again there will be a deadlock between those who think the conceptual link between know-that and articulation has priority, and those who think that the connection between know-how and skilled action does. Conceptual analysis does not provide us with the means to settle this, even with semantic considerations thrown in. This again leaves conceptual analysis ill-equipped to provide an interesting conclusion. It does, admittedly, tell us that intellectualism, by the letter, is true but it is unable to tell us whether or not skilled action involves know-that.

This suggests it would be worth looking to an alternative strategy in order to investigate intellectualism in more depth.

#### **4. Psychologism**

According to psychologism, intellectualism is an empirical thesis to be decided by investigation into our psychological structure. Know-how and know-that are

psychological states, and thus it is for psychology to determine whether the former is an instance of the latter. This gives us the following view:

*Psychologism*: Find the appropriate psychological kinds to be identified with know-how and know-that, then investigate (empirically) the relationship between the two kinds of state.

Psychologism seems to avoid the problems conceptual analysis has in making progress on the nature of intellectualism. Looking to our best scientific theory of the mind is a sure-fire way to obtain interesting results relevant to both philosophy of mind and epistemology. It's also not going to run into the kind of *impasse* that irreconcilable intuitions create for conceptual analysis, since we're dealing with testable hypotheses about an objective subject matter. Note that a key issue here is what makes it the case that a psychological kind is an *appropriate* candidate to be identified with a given type of knowledge: reflection on this leads to the problems I'll now discuss.

#### 4.1 Problems with Psychologism

A popular thought is that know-that is to be identified with *declarative knowledge* (a putative psychological kind) and know-how with *procedural knowledge* (another putative psychological kind).<sup>19</sup> Intellectualism then amounts to the claim that all procedural knowledge is declarative knowledge – an empirical thesis. Of course, this proposed identification must be justified

First, we must specify conditions for when a psychological kind is fit for identification with a folk concept such as know-how or know-that. As Devitt notes, many cognitive scientists make the above identification, so one might be tempted to think we should defer to the experts here – but this would be a mistake.<sup>20</sup> It is clear that cognitive scientists can be pretty loose with the term 'knowledge': they are rarely concerned with whether the states in questions meet a justification condition, and in some cases don't

---

<sup>19</sup> This version of psychologism is defended by Devitt (2011), Adams (2009) and Wallis (2009).

<sup>20</sup> See Devitt (2011). However, Stanley points to instances in which psychologists reject these identifications – see his (2011) ch 7.

even care whether it's factive.<sup>21</sup> By analogy, we shouldn't conclude that fairly simple computers have mental states from the fact that computer scientists talk about what such systems 'know' – theorists will often use such terms *opportunistically*, to allow readers to get a rough intuitive grip on their ideas. This suggests we should have some independent means to check whether their use constitutes a genuine reduction.

We need to show that the relevant psychological kind possesses features that make it a suitable candidate for knowledge. The standard naturalist line here is to look for the natural kind which roughly fits with the folk application conditions. Therefore, we need a psychological kind that is present in a good number of paradigm cases of knowledge, and that satisfies some of the central intuitive principles about knowledge states.<sup>22</sup> Crucially, these criteria of fit must be loose enough to remain neutral on the contested intuitions discussed in the previous section; otherwise we will fail to avoid the *impasse* that upset conceptual analysis.

In a best-case scenario, there will be a *unique* psychological kind meeting the uncontentious criteria for know-that. One will be able to identify this state with know-that without begging any conceptual questions, and then go on to investigate whether such a state is present in typical cases of know-how.

If there are too many kinds in the ballpark, however, psychologism will be unable to provide a candidate unambiguously. Take my knowledge that Morpeth is in the north of England. As mentioned, I can both assert this when asked and use the information to guide action where necessary. So I have a *token* knowledge state which is an instance of both a restrictive type that entails assertibility, and a broader type that only requires action guidance. Now if only one out of the two state-types is a psychologically significant kind, then psychologism tells us that this state is know-that. And we have a verdict on intellectualism one way or the other, depending on whether it's the narrow or broad state. However, if *both* are psychologically significant kinds, psychologism will not deliver a verdict as to which of the two is know-that – since both fit roughly with our ordinary concept of knowledge. Thus the plausibility of psychologism hinges on whether the uniqueness condition is met.

---

<sup>21</sup> See Machery (2009) p 8.

<sup>22</sup> See, e.g., Kornblith (2002), for this kind of approach.

Let's look at how this relates to the psychologist's claim that know-that should be identified with declarative knowledge and know-how should be identified with procedural knowledge. To see whether this is acceptable, we must get clear on what the declarative/procedural distinction amounts to. As Stanley (2011) argues, it is not entirely clear how this distinction is to be understood in psychology – as opposed to computer science where it originated (where it distinguishes information explicitly encoded in the computer's memory as opposed to information 'contained in the program').<sup>23</sup> A first pass attempt at this would be to identify states of know-that with *sentential* representations – that is, sentences in the *language of thought*.<sup>24</sup> This seems like a reasonable gloss on what it is for a state to be declarative rather than merely procedural.

Of course, not just any sentential representation is fit to be identified with knowledge. A basic point is that not all sentential representations can be knowledge states since some of them will be *false* and others may function as desires rather than as *information*. So to get something plausible we need to add in the assumption that there is a psychologically significant distinction between those sentential representations that are information states and those that are desires (or any other attitude). Second, we need to assume that there is a psychological distinction between information states that are true and (appropriately) justified and those that are not – i.e. between knowledge and *mere belief*. The best bet for this would be to single out representations formed as a result of non-defective (and so reliable) processes.<sup>25</sup>

An old objection to this kind of view is that it holds our account of knowledge hostage to empirical assumptions about the nature of the mind. If it turns out there is no language of thought, it seems that there will be no knowledge states.<sup>26</sup> However, given the problems with conceptual analysis, this argument should not impress us. We shouldn't expect to be able come up with conditions for knowledge that are *a priori* secure. Instead, it's enough that our account be based on premises that are good working assumptions, given our current empirical knowledge. It's widely held that the claim that our cognitive

---

<sup>23</sup> Indeed, one might be sceptical this distinction can be meaningfully applied in psychology at all – this of course would only bolster my case that we should seek an alternative methodology.

<sup>24</sup> And thus follow in the footsteps of the traditional psychologicistic account of *belief* proposed by Fodor (1975).

<sup>25</sup> If this strategy fails, it's natural to take psychologism to entail that there's no such thing as knowledge, only belief.

<sup>26</sup> See, e.g., Lewis (1994) or Jackson and Pettit (1990) for this kind of argument.

system involves sentential representations *is* a good working assumption – and I won't dispute this.

However, a sentential account is in danger of getting the scope of know-that radically wrong. All sorts of clearly sub-personal states quite possibly involve sentential representations; for example, the unconscious processing that calculates what sentence has been uttered based on the sounds one hears.<sup>27</sup>

Fortunately, there is a psychologically significant distinction that promises to rule out such clearly sub-personal states: that between representations in *modules* and representations in the *central system*.<sup>28</sup> Thus, I think the best bet for a psychological state that lines up with know-that is the following:

*Psychological Knowledge*: S knows that p iff p is true, and S has a centrally stored sentential representation with p as its content such that the representation is non-defectively formed.

This account is not conceptually question begging, since it doesn't build-in an articulation condition; it's a substantive empirical question whether we have conscious access to all centrally stored sentential representations. Further, this is the reduction Devitt seems to lean towards, which suggests this account is charitable to the psychologist.<sup>29</sup>

If such an account were empirically secure – i.e. if its commitments were plausible working assumptions – and fit roughly with our intuitive application of knowledge, then psychologism would be a promising approach. However, this is not the case. There are a number of reasons to doubt the assumptions built into *Psychological Knowledge*.

The first problem concerns the notion of a *central system*. Though it seems safe to assume that at least some mental processes involve sentential representations, the same cannot be said for the assumption that there is a central system. On the one hand, there have been many challenges concerning the extent to which perceptual systems do in fact constitute modules. As Prinz notes, there is a great deal of evidence suggesting there can

---

<sup>27</sup> See Stich (1978).

<sup>28</sup> See Fodor (1983).

<sup>29</sup> See Devitt (2011) p 210.

be top-down influence on our perceptual states – for example when listening to a sentence with a deleted phoneme, we will hear the sound that makes for a coherent sentence. He notes that “If subjects hear, ‘The ‘eel is on the axel,’ they experience a ‘w’ sound in the gap. If they hear, ‘The ‘eel is on the orange,’ they experience a ‘p’ sound.”<sup>30</sup>

On the other hand, many authors reject the idea of a single central system and suggest the mind is ‘massively modular’ so that even the representations involved in the various conscious processes do not constitute a single integrated system.<sup>31</sup> Pinker (1997) argues that the mind has separate modules dedicated to, among other things, humour, mindreading and sexuality. Together, this suggests that it is at the very least an open question whether psychology will present a clear cut distinction between central and non-central representational states, fit to mark the boundaries of know-that.

Another worry comes from so-called ‘junk sentences’ in the long-term memory. It’s plausible to think that a whole lot of information gets stored in our long-term memory that is almost never employed. Indeed, Mandelbaum (2014) argues that every proposition we ever *entertain* is stored in long-term memory. The psychology of judgments and decision-making investigates whether and when various bits of information are employed in guiding action.<sup>32</sup> The key point for our purposes is that a sentential representation (even one in the central system) must stand in the appropriate relations to the mind’s heuristic machinery to be of any significance. It’s not clear that there is a psychologically significant distinction between those representations apt to get called up in a wide range of situations and those which are generally idle.<sup>33</sup> This suggests that *Psychological Knowledge* would require radically expanding the extension of knowledge, to include all sorts of things we have no interest in.

A different set of empirical considerations arise from the fact that there are quite possibly many mental states which are not *sentential* representations, yet seem like instances of knowledge. One example of this is visual memory. Often such mental states guide our behaviour – they are both consciously accessible and have an action-guiding role. However, whether mental images involve purely sentential representation is an open

---

<sup>30</sup> Prinz (2006) p. 31.

<sup>31</sup> See Pinker (1997), Carruthers (2006).

<sup>32</sup> See, e.g., Payne and Bettman (2004).

<sup>33</sup> See Mandelbaum (2014) for an argument that all stored sentences are of a kind, from the perspective of psychological theory.

question. It is a claim rejected categorically by Kosslyn (1996), and in more moderate fashion by Tye (2000).<sup>34</sup>

A second example is that of *analogue magnitude states*. Beck (2012) has argued that a distinct type of representation is formed when we have to make a snap judgment about the magnitude (number, length etc.) of some object(s) – when we are not able to count or measure. In his words we form a ‘noisy, analogue’ representation; one that allows us to make rough judgments but not fine-grained ones. If Beck is right, such states cannot be sentential representations but we might still want to classify them as knowledge. Therefore, even if there is a respectable psychological kind lining up with the notion of a sentential representation in the central system, it is quite possible that such a kind will exclude much of what we want to count as knowledge.<sup>35</sup>

This shows that it is far from a safe bet that *Psychological Knowledge* gets us a state that fits comfortably with our concept of knowledge. Moreover, I think the discussion also shows that no amendment will be able to rectify this. It’s plausible to think that our best taxonomy of psychological states will be very fine-grained. First, this will include different kinds of analogue and sentential states. Second, such states can be classified according to which module they belong to – whether these are traditional perceptual modules or the central modules entertained by Pinker. Third, states can be classified in terms of how likely they are to be *accessed* in different circumstances, in virtue of the agent’s structure of heuristics. Any single one of the categories is going to be too narrow to identify with knowledge. Moreover, there doesn’t appear to be a single theoretically privileged way of grouping together a certain collection of these fine-grained states as different kinds of know-that – there is no theoretically privileged place to draw the line.

As I mentioned at the beginning of this section, the hard-core psychologist might at this point argue that the correct conclusion to draw is an error theory about knowledge. Either, we should go with *Psychological Knowledge* despite the radical consequences, or

---

<sup>34</sup> Though Pylyshyn (1981) defends a sentential account of mental imagery.

<sup>35</sup> Another potential problem for *Psychological Knowledge* is Kornblith’s (2002) argument that in *cognitive ethology*, the broad notion of a reliable information state is the theoretically significant kind. I won’t discuss this in detail here because it would involve digressing into a debate about which branches of science are relevant for *Psychologism*.

replace our single concept of knowledge with a family of more fine-grained notions – or perhaps we should scrap talk of knowledge entirely!<sup>36</sup>

I've been criticizing extreme versions of conceptual analysis and psychologism, but I think the discussion also suggests that an approach that takes elements of both – one that recommends some system for looking at a mixture of intuitions and empirical results to determine the nature of our concepts – looks no better. We saw how conceptual analysis struggled to decide between articulation-first intuitions and action-first intuitions; moreover, the psychological evidence presents us with a great range of natural kinds, with no theoretical reason to pick out on favoured subset as knowledge. There doesn't seem much reason to think combining the two would help matters.

## 5. An Alternative Strategy

My suggestion is that we should look to the practical role of knowledge attribution to gain insight on the topic of intellectualism. Our everyday practice of knowledge attribution – and attitude attribution more generally – is incredibly useful. It seems unlikely that this utility would be preserved if we took the radical revisions recommended by psychologism. Investigating why it is useful and what account of knowledge fits with its practical role is a matter of philosophical interest – one that goes beyond studying ordinary language intuitions for the sake of it. *This* is why I think the revisionary conclusions of psychologism should be resisted.

In order for this to be a viable strategy, pragmatism will have to provide interesting results as to the nature of knowledge. I will say a little now about why we can expect such results, before getting into the details of the account in the next chapter. The version of pragmatism I will be working with consists of a three-step process:

1. Identify the practical role of knowledge attribution.

---

<sup>36</sup> This is what Churchland (1981) proposes in the case of belief. Another line of response is for the psychologist to appeal to further assumptions about the nature of reference to settle the matter. It might be that a richer naturalistic theory – such as Fodor's (1990) 'asymmetric-nomic-dependence' account – provides a principled way to settle the matter by the psychologist's lights. Again, though, this looks to take us so far away from the practically interesting notion of knowledge and the other attitudes, that it's worth looking at an alternative approach. Thanks to Ted Bach for pointing out this line of argument.

2. Find the state that best fits with this practical role.
3. Identify this state with knowledge.

In recent work, philosophers have argued that this approach can resolve a range of long standing epistemological disputes.<sup>37</sup> As was mentioned above, pragmatism is not a suitable strategy for theorizing about *all* our concepts – only some of them have a practical role worth investigating. We need to show that knowledge, in particular, is well suited to this process.

We use many terms to refer to a particular kind of thing, and doing so can be useful for all sort reasons in various situations, but there is nothing particularly interesting to say about the practical role of the term. Take the term ‘magnesium’, for example: we use it to pick out the chemical element *magnesium*, and that’s about all that can be said at a general level.

Other terms pick out entities that have a particular *function*, while having little else in common. For example, ‘knife’ picks out, roughly, the class of entities that can cut through various materials in a characteristic manner. Here, there is something a little more substantive to say about the practical role of the term – it picks out an entity that can perform a particular function, because we have an interest in taking advantage of that function.

Finally, though, we have terms which do not pick out obviously functional kinds but that do seem to have an interesting practical role. Take a term like ‘tasty’ – it doesn’t pick out a natural kind, or a class of entities that perform a particular function, but it still seems *useful*. Investigating why looks like a fruitful project that will shed light on the nature of tastiness. There are a number of concepts which seem to fit this pattern – e.g. *truth*, *goodness*, and *number* – and a range of philosophers have investigated their practical role.<sup>38</sup>

---

<sup>37</sup> Most famously Craig (1990), but also Dogramaci, (2012), Hannon (2013), Henderson (2009), and Pritchard (2012).

<sup>38</sup> As Dogramaci (2012) notes, Quine’s (1970) motivation for deflationism about truth, is based on an insight about what makes the concept *true* useful. Underlying Gibbard’s (1990) theory of ethical concepts is a picture of why ascribing praise and blame is useful. Integral to Yablo’s (2005) fictionalist account of numbers is a view of why it’s useful for us to quantify over numbers.

I aim to show that knowledge falls into this final category. It should be obvious it does not pick out a kind with a straightforward function – like ‘knife’ does. It also doesn’t look like it has an uninteresting role, like ‘magnesium’. Our criticism of psychologism shows it’s unlikely to refer to a natural kind. However, as I’ll discuss in the next chapter, it is useful in allowing us to predict and explain behaviour (as many other philosophers have observed). Further, communicating what someone does and doesn’t know is something we attach great significance to – knowledge is in some sense *praiseworthy* while lacking knowledge is in some sense *deficient*. This provides further evidence for practical significance.

Before moving on, I want to make clear that pragmatism is *not* simply a modified version of conceptual analysis. I take conceptual analysis, as a matter of stipulative definition, to be a project carried out from the armchair – canonically, it concerns only *a priori* matters. Analysis of the practical role of a term, however, is an *empirical* matter – it involves looking into how a term is used in a social setting. Though proponents of a priori conceptual analysis sometimes assume they can appeal to practical role, they are mistaken. The point is well put by Block and Stalnaker:

Jackson emphasizes that there is an element of stipulation involved in the a priori entailments from physics of everything outside of consciousness. [Jackson says the analysis must be ‘near enough’ to everyday application.] Near enough for what? His answer is: near enough for practical purposes... But it can hardly be assumed that what matters for practical purposes is a priori. What matters for practical purposes depends on the facts of psychology and economics (for example). [Block and Stalnaker (1999) p. 27]<sup>39</sup>

This insight is central to my understanding of pragmatism: we must look at the role of a term in a *social setting*, and not simply rely on our intuitions about the matter. This is important since the term ‘pragmatism’ is applied in many different ways – for example, Fodor (2008) equates it with an inferentialist notion of meaning, which is a separate

---

<sup>39</sup> Though Chalmers and Jackson (2001) offer a response to Block and Stalnaker in defence of conceptual analysis, they unfortunately do not engage with this particular issue.

matter entirely and rather hospitable to conceptual analysis. It is the relevance of these facts about a term's social role which allows pragmatism in my sense to gain fresh traction when conceptual analysis stalls.

## 6. Reference and Practical Interests

Even if one grants that the practical role of terms like 'knowledge' or 'tasty' are of interest, one still might doubt they can shed light on the concepts themselves – that is, they might resist step 3 of the pragmatist approach. A sceptic might phrase what's going on with maximum uncharity as follows: "You are arguing that since it would be nice and helpful if 'knowledge' referred to state XYZ, 'knowledge' *in fact* refers to XYZ."

Putting aside this rhetoric, I do think we have good reason to endorse the proposed relationship between practical role and meaning, as I'll now explain. I can't argue for this conclusively since it would require getting to the bottom of the nature of reference – a large digression indeed. Instead, I propose to show how practical role could enter into meaning, if we accept a particular theory of reference – Lewis' eligibility theory. Even if one does not like the background theory, this shows the possibility of a pragmatic element to reference, by providing proof of concept.

### 6.1 Lewis' View

The background theory I will be working with is the theory of reference laid out by Lewis (1984) in his paper 'Putnam's Paradox'.<sup>40</sup> The view proposes a way to provide an *interpretation* of a language. An interpretation is a function that assigns objects to names and properties to predicates, as their referents (for simplicity we'll ignore other elements of the language). The Lewisian theory states that the *correct* interpretation of a given language is the one that best meets two constraints: *charity* and *eligibility*. I'll explain what these constraints amount to in turn.

---

<sup>40</sup> Although controversial, a number of philosophers have influentially defended it – e.g. Weatherson (2003) and Sider (2013).

Lewis claims that the folk's use of a given language determines a *folk theory*. The folk theory consists of a set of sentences that are in some sense obviously true. These are sometimes referred to as 'the platitudes'. There is some disagreement as to what precisely determines the folk theory, but this need not concern us: the key point for our purposes is that whatever the platitudes are, they can be discovered by conceptual analysis, and so they are not determined by the *practical role* of the terms. Lewis took investigating the platitudes to be an armchair project – and we saw above that one cannot discover the practical role of terms in this way. The notion of a folk theory allows us to define charity: an interpretation is *charitable* to the extent that it makes the folk theory true.

In order to define eligibility, Lewis appeals the notion of *perfectly natural* properties: these are the properties involved in the laws of (correct) fundamental physics. He also uses the notion of *relative* naturalness, which is derived from the perfectly natural properties: the 'simplicity' with which a property can be constructed out of the perfectly natural properties determines its level of naturalness.<sup>41</sup> An interpretation is *eligible* to the extent that it assigns natural referents to the elements of the language.

Therefore, on Lewis' view, the correct interpretation of a language is the one that, as well as possible, both makes the platitudes come out true and assigns natural referents to names and predicates. This means that both conventions of language use and the underlying nature of the world play a role in determining reference. Lewis is making a compromise between traditional descriptivist theories of language on which it is only the inferential connections between terms that determine meaning, and naturalist theories on which reference is determined solely by empirical scientific matters.<sup>42</sup>

Indeed, one can move towards one end of the spectrum or the other depending on how one weights the two constraints. On a view that prioritizes charity, overwhelming

---

<sup>41</sup> How to define the relevant kind of simplicity is a tricky business – see Williams (2014) for a discussion. On top of this notion of relative naturalness, one might be tempted to appeal to *emergent* higher-order naturalness as well. I want to put this aside because there's the potential for the practical significance of a property to ground a certain kind of higher order naturalness. This would muddy the waters when trying to decide how practical role enters into a theory of reference.

<sup>42</sup> My statement of the Lewisian theory is brief and informal. To better understand the motivation for the view requires looking to the problems with previous theories that Lewis was addressing, as well as Lewis' broader theory of mind, language, and metaphysics. A good place to start with this is ch. 1 of his (1986). Further, I have glossed over many technicalities that are necessary to give an adequate formal account of an interpretation of a language, and the way in which the correct interpretation is selected. Williams (2007) lays out the necessary details with great care. These details do not, though, affect the force of my argument.

considerations of naturalness are required to justify an interpretation on which some small element of the folk theory is false. This gives us something close to traditional descriptivism. However, a view that prioritizes eligibility – so that terms always refer to the most natural kind available so long as it allows for some basic elements of the folk theory to be true – gives us a pretty strong naturalism about language.

Despite this flexibility, I think we need to add a third constraint to Lewis' theory if it is to be tenable: *practicality*. This states that the correct interpretation of a language should make it *useful*. If the folk's use of a term achieves a certain purpose, an interpretation should assign a referent that fits with that purpose. The correct application conditions for the term should allow it to serve its purpose optimally. For example, if our use of the term 'tasty' serves the purpose of drawing our attention to foods that evoke a pleasurable gustatory sensation, the term should refer to a property that is possessed by the foods that are apt to produce such a sensation.

To see why this constraint is needed, we need to look at some problem cases for Lewis.

## 6.2 The Strategy

Problems for the Lewisian theory are most apparent in cases where *charity* and *eligibility* come into conflict. This happens when a term is used in such a way that there are two candidate referents, one natural and the other unnatural: if it refers to the natural one, some platitudes will be false; if it refers to the unnatural one, these platitudes will be true (call this a case of *interpretation conflict*). To decide what the term refers to, the Lewisian must look at how *charity* and *eligibility* are to be weighed against each other. A fully spelled out version of the theory must specify a way of assigning weight to the level of charity and level of eligibility achieved by an interpretation, in order to rank them – and thus decide which is the correct interpretation. Lewis does not need to precisely specify the weighting method himself; however, he is committed to there *existing* some weighting method that assigns the correct referents to all terms in cases of interpretation conflict.

I aim to show that without a practicality constraint, no such weighting method can exist. My strategy is as follows:

- 1) Identify a pair of cases of interpretation conflict where the trade-offs between fit with intuition and naturalness equivalent.
- 2) Argue that the reference of the terms differ in these two cases.
- 3) Show that there is a practical difference between the two cases.
- 4) Conclude that practical matters play a role in determining reference.

If stages 1) and 2) go through, then Lewis' theory as it stands cannot be correct. If it's possible for facts about charity and eligibility to be the same when facts about reference are different, then one cannot explain the latter solely in terms of the former. If it can also be shown that 3) holds – and indeed that it holds in a range of cases – then inference to the best explanation favours my conclusion.

### **6.3 Defective Folk Theories**

The cases of interpretation conflict I will be looking at involve *defective folk theories*. These are cases in which subjects have vocabulary describing a certain aspect of the world, but are ignorant of the underlying facts. They apply this vocabulary in such a way that *if* their terms referred to the natural kinds, some of the platitudes would be false. The natural question is what the terms in fact refer to in such cases. There are three kinds of answer here, each of which applies to various historical examples.

The first is *Reduction* – this is when the terms refer to the natural kinds and some platitudes are false. This involves positing a kind of error theory. An example of this concerns various ancient cultures' use of the term 'star'.<sup>43</sup> They applied this term to some celestial objects but not others. In particular, they did not apply it to the sun, but did apply it to the other stars. For them 'the sun is not a star' was a platitude. However, the property satisfied by all the stars, including the sun, is a natural one, since they all share important

---

<sup>43</sup> Here, and in all the examples that follow, I ignore the complication that the subjects were not speaking modern English – one can view the English words as stand ins for the appropriate translations.

physical characteristics, while the property satisfied by all stars except the sun is quite unnatural. To describe the situation more naturally: the ancients said the sun was not a star but they were mistaken. Therefore, the ancients' term 'star' picked out the property *star* and their folk theory was mistaken.<sup>44</sup>

The second possibility is *Disjunctivism* – this is when the terms refer to some highly unnatural property in such a way that the platitudes come out true. An example of this concerns use of the term 'jade' prior to the mid 19<sup>th</sup> century. The term 'jade' at that time was applied to certain green precious stones – for various exemplars 'this is jade' was a platitude. However, in 1863, it was discovered that the exemplars of 'jade' came in two different varieties: the chemical compounds *jadeite* and *nephrite*. In this case, disjunctivism holds: before (and after) 1863, the term 'jade' picked out the disjunctive property that applies to both jadeite and nephrite rocks. When people described these rocks as 'jade' they stated a truth.

The final possibility is *Eliminativism*: this is when terms in the defective theory are empty – they do not refer. An example of this is found with the phlogiston theory of heat. This was a 17<sup>th</sup> century scientific theory that described the processes surrounding combustion using a family of terms like 'phlogisticated' and 'dephlogisticated'. The theory was radically mistaken since it took things like the following to be platitudes: 'phlogiston is a substance present in all combustible bodies, and that is released during combustion'. The correct thing to say here is that there is no such thing as phlogiston – when 17<sup>th</sup> century scientists talked about 'phlogiston' they weren't referring to anything at all.

According to the Lewisian theory, it is the different ways in which charity and eligibility are more or less satisfied that determines the correct interpretation of a defective folk theory. If the gains in naturalness outweigh the violations of folk theory, then *Reductionism* obtains – if it's the other way around then *Disjunctivism* obtains. If nothing does well enough with either then *Eliminativism* holds. This could be because it's built into the folk theory that the referents must be natural in some way and no such potential referents exist – meaning that eliminativism undermines the folk theory less than disjunctivism.

---

<sup>44</sup> This example, and its interpretation in a Lewisian context comes from Weatherson (2003)

## 6.4 The Problem Pair

Here are the cases that present a problem for the Lewisian. First, look again at the phlogiston theory, and in particular consider why disjunctivism is not applicable in this case. As Churchland (1981) noted, one could find unnatural kinds such that if the concepts of the theory referred to them, the bulk of the theory would end up coming out as true. In his words “A cracking good defence of the phlogiston theory of combustion can also be constructed along these lines. Construe being highly phlogisticated and being dephlogisticated as functional states defined by certain syndromes of causal dispositions”<sup>45</sup> – these dispositions would presumably involve things like a tendency to react to flame in a characteristic manner. Such functional kinds could be constructed so as to uphold a good number of the platitudes of the phlogiston theory, things like ‘highly phlogisticated objects burn for a long time when put in a fire’.

The Lewisian could argue that not enough of the folk theory is upheld and too much unnaturalness is admitted for this to be preferable. One could also point to the fact that the claim ‘phlogiston is a *substance*’ is a platitude, which suggests that it’s part of the folk theory that the term is a natural kind – having it refer to a functional kind violates charity even more than eliminativism does.

The problem is that the Lewisian then seems backed into a corner when it comes to an analogous example. I have in mind the ancient elemental theory. The ancient Greeks had a theory of the world according to which every material object was made up of some combination of earth, air, fire and water (and according to some a fifth element known as ‘aether’). This theory was radically wrong, obviously: first there are far more than four elements; second the exemplary cases described as ‘earth’, ‘air’, ‘fire’ and ‘water’ are not elements at all. In light of this, we must ask what the terms referred to. The right approach for ‘earth’ and ‘air’ is disjunctivism – they referred to earth and air respectively; for ‘water’ and ‘fire’ it’s reductionism since water and fire are fairly natural kinds. We say, ‘the ancients said some very weird and totally mistaken things about earth and air’ – which implies that their term ‘earth’ referred to the disjunctive kind *earth*.

---

<sup>45</sup> Churchland (1981) p. 80.

I don't think, though, that the Lewisian can explain this result. She needs to identify the difference between 'earth' and 'phlogiston', which explains why disjunctivism applies in the former case, and eliminativism in the latter. There seems no good reason to think that the functional kind which could have saved phlogiston is significantly less natural than *earth*. Further, it does not seem plausible that it was less of a violation of folk theory to allow disjunctivism for the elemental theory, than it would have been in the case of phlogiston. It's not plausible to think that the elemental terms were intended to pick out functional kinds – or even that the folk theory allowed for it. They were part of Aristotle's *Physics*, which seems to put them on a par with the concepts of phlogiston theory. Indeed, the elemental theory was fundamental to the ancients' view of the world: to give it up would have upended pretty much their *total* folk theory.

What I think is the key difference between the two cases is that it's *useful* to refer to earth and air, even if they're not natural kinds. We have an interest in talking about the 'stuff on the ground' and the 'stuff we breath in' in our everyday lives. On the other hand, if the phlogiston theory proves false, there's no real benefit to being able to talk about some weird functional properties that partially fit with the tenets of the theory.

## 6.5 Further Cases

The Lewisian might resist by pointing to another difference between the cases. Phlogiston theory was a full-fledged scientific theory, outside of folk-discourse. On the other hand, though the elemental theory was employed in a scientific manner, the terms for the elements were also used in everyday contexts. This, it might be argued, makes disjunctivism less of a violation of charity in the case of elemental theory than in the case of phlogiston theory.

I have two points in regard to this response. First, the Lewisian must motivate the claim that the difference she points to is in fact a good *explanation* of the results. Otherwise, it seems she is simply finding a difference between the cases and then adding it to her theory of reference *ad hoc* to secure the desired results. To me, the explanation of the difference that appeals to the practicality constraint is much more intuitively satisfying.

A second point is that since we need to look at the practical role of language use in non-contrived settings, the different examples we consider are always going to be highly complex. This means we'll never be able to find the perfect minimal pair, and so the Lewisian will always have a certain degree of wiggle room, appealing to subtle differences between the scenarios. Unfortunately, this means my argument for the practicality constraint is necessarily an inference to the best explanation rather than a logical entailment. However, I will present a number of additional examples to bolster my case.

Consider the example of superstitious talk of 'spirits' and 'ghosts'. As sensible people, we know that there exist no supernatural entities for these terms to refer to. The correct interpretation here is eliminativism: ghosts and spirits do not exist. We could, though, interpret these terms as picking out unexplained noises and movements, strange winds and physical sensations etc. This would make them disjunctive terms but would vindicate much of folk usage.

In this case, these terms are not part of a fully-fledged scientific discourse and are deeply engrained in everyday use – especially if we look at how they were used in earlier times. This means that the feature the Lewisian had relied on to motivate eliminativism in the case of phlogiston theory is not available here. Thus the question remains: why does disjunctivism apply for 'earth' but not 'ghost'? Again, I would point to practical considerations. We have no particular use for referring to the category of strange noises &c. if they are not related to supernatural phenomena. Therefore, the practicality constraint motivates eliminativism.

This discussion outlines why adding a practicality constraint to one's account of reference is appealing, to me at least. It is far from a decisive argument, however. The best way to make the case, I think, is to look at as many examples as possible. One case I find particularly compelling is Mark Wilson's (2014) discussion of temperature. The claim that 'temperature' picks out the property of mean kinetic energy is seen as a canonical case of reduction, motivating naturalist theories of reference. However, the details complicate things somewhat. The famous identity statement is only true for gases. Liquids and solids have what's known as a 'frozen order' which makes it impossible to

get a sensible measurement of their mean kinetic energy.<sup>46</sup> There are ways to correct for this by moving away from the familiar equation. For example, in rubber bands, we get the right result by finding ‘higher scale ensembles of possible wiggling of molecules’ to form ‘quasi equilibrium.’ This allows us to obtain a reading of temperature which matches the behaviour of the material in the appropriate way. Further, even more complicated things must be done with other materials to get a reasonable concept of temperature.

Therefore, though the natural kinds underlying our application of the term ‘temperature’ play a role in determining its reference, they are not sufficient to explain things completely. The term in fact picks out a relatively disjunctive category and this is motivated by what, practically speaking, is useful when carrying out scientific investigation. In his book *Wandering Significance*, Wilson (2006) presents a vast range of examples describing the complex relationship between terms and their referents, and the essential role played by practical considerations. Situations like the one I described for temperature apply to *weight*, *solidity*, *hardness*, and many more.

This completes my case for a practicality constraint on reference. If the Lewisian theory is correct, the correct interpretation of our language is determined by appropriately balancing *charity*, *eligibility*, and *practicality*. To reiterate, I am not endorsing the Lewisian assumptions, but showing how they require a pragmatic constraint on reference to be added.<sup>47</sup> This is enough, I think, to justify exploring my claim that when coming up with an account of the attitudes, one should look for a state that fits with the practical role of belief and knowledge ascription.

---

<sup>46</sup> Wilson (2014) provides a more detailed explanation of this, as well as references to the scientific literature.

<sup>47</sup> I hope, though, that it can be seen how my argument could be tweaked to apply to competing theories of reference.

## Chapter 2

# The Practical Role of Knowledge Attribution

### 1. Introduction

In the previous chapter, I argued that we should pursue pragmatism in order to better understand the nature of knowledge. In this chapter, I will investigate the practical role of knowledge attribution and its consequences for epistemology.

Most recent work has followed the lead of Edward Craig in his book *Knowledge and the State of Nature*, who links the practical role of knowledge attribution to the provision of testimony: we use the term ‘know’ (and its cognates) to pick out *good informants*. I will argue that this emphasis on testimony is mistaken, and present an alternative picture on which the practical role of knowledge is connected primarily with *action*. Roughly, the term ‘know’ is used to pick out someone who will successfully complete their tasks in a relevant area.

Deciding between the testifier account and the action account has implications for epistemology more generally – showing that the debate is substantive. For while the testifier account fits naturally with an internalist conception of knowledge (prioritizing conscious access and the provision of reasons), the action account pushes us towards an account of knowledge linked more to action guidance than conscious deliberation – and thus to reliabilism over internalism. Importantly, the debate also sheds light on the topic of the first chapter: the testifier account motivates an anti-intellectualist account of know-how, while the action account motivates intellectualism.

A modest conclusion, that I hope a range of readers are convinced by, is that deciding between the testifier account and the action account is of general significance in epistemology. It is important for those interested in the epistemological issues I discuss to investigate the practical role of knowledge attribution. A more ambitious and contentious

conclusion I argue for is that the action account is correct, and that we should accept the epistemic consequences it motivates. I hope that those who are not persuaded to endorse the ambitious conclusion will still take the modest conclusion seriously.

## 2. Framing the Debate

We discussed in the previous chapter initial reasons why the practical role of knowledge attribution is worth investigating: it doesn't seem like any psychological kind lines up with our concept and yet it's something we attach great significance to, and seems useful in our everyday practice. Now, I propose we look at what the practical role might be.

### 2.1 A Minimal Proposal

A natural starting point for pragmatism is the familiar idea that the folk psychological attitudes in general are useful because they allow us to predict and explain behaviour. Moreover, they allow us to do so in an *efficient* manner. Roughly speaking, agents act in ways that *make sense* given their beliefs and desires – that is, they act in ways that would tend to satisfy their desires, were their beliefs true. Such a formula is both systematic, and relatively easy to use.

This second point is crucial since, as Dennett has long argued, if one wants to predict behaviour as accurately as possible with no constraints on the calculations required to make the predictions, one's best bet would be to take a complete description of a subject's microphysical structure and the structure of her environment, then deduce her sequence of behaviour from the laws of physics.<sup>1</sup> However, this is obviously not something we could actually do. What attitude ascriptions allow us to do is predict behaviour reasonably well in everyday situations – employing only everyday evidence (not MRI scans) and relying on only fairly simple reasoning.<sup>2</sup>

We cannot stop here, since even if this is a good account of the general practical role of attitude ascription, it's not clear how well suited it is to explain *knowledge* ascription

---

<sup>1</sup> See Dennett (1989) – especially his discussion of the super smart Martians on pp. 26-27.

<sup>2</sup> This highlights an additional problem with replacing folk psychological concepts with those of psychology. Though psychology gives a somewhat more compact description of agents than physics does, it's still far too complex to be used in everyday situations.

in particular. A natural thought is that in order to predict someone's behaviour one simply needs to know what they *believe*, not what they know. For example, if I'm thirsty and I believe there is a glass of water in front of me, you can predict I'll reach for the glass – whether my belief amounts to knowledge or not. This line of argument has been contested, most notably by Williamson (2000). He uses his notorious example of the diamond thief to make the case that often knowledge attributions better explain behaviour than mere belief attributions. I am sympathetic to Williamson's position, though I will be approaching it from a different angle. I believe that the explanatory power of knowledge attributions is *practical* in nature, and so looking at the practical role of knowledge attribution sheds light on knowledge-based explanations.<sup>3</sup>

Thus, one challenge is to explain the practical role of knowledge attribution, as opposed to mere belief attribution. A particularly important feature here is the *normative* nature of knowledge – in the evaluative sense of the term. We view knowledge as a *good thing*; and beliefs that fail to amount to knowledge are, in some sense, defective and worthy of criticism.<sup>4</sup> If the purpose of knowledge were simply to predict behaviour, this would be utterly mysterious. We seem to have an interest in *promoting* knowledge over mere belief, and this is something pragmatism should be able to help make sense of.<sup>5</sup>

A final point is that we are looking for an elaboration of pragmatism that is detailed enough to allow us to draw interesting conclusions about the nature of knowledge. 'Predicting and explaining behaviour' is a very broad and vague task, and so seems ill-equipped to handle this. If we were interested in predicting verbal behaviour and explicit reasoning, we would look for a testimony-based account, while if we were interested in non-verbal behaviour, we would look for an action-based account – so more detail on the practical role of knowledge is needed.

## 2.2 Clarifying the Project

---

<sup>3</sup> I am not committed to the claim that knowledge attributions are well suited to *scientific explanations*. Williamson at times seems to suggest that cognitive science ought to work with knowledge rather than belief – see especially his (2006). I don't want to say anything so bold.

<sup>4</sup> Though there may well be a mitigating explanation as to why the agent holds such beliefs.

<sup>5</sup> Whether it gives the whole story will depend on whether there is a deeper notion of epistemic value which goes beyond the pragmatic. I discuss the relationship between pragmatism and the normativity of rationality in ch. 5.

My criticism of the minimal account provides us with three criteria for an adequate pragmatist account:

- C1. Explain the practical role of knowledge attribution in particular.
- C2. Explain the normative element in knowledge attribution.
- C3. Provide a specific enough account to deliver interesting epistemological conclusions.

The first two criteria make sure that the account of the practical role is on the right track. The third ensures that the account has *bite* – and is of wider significance for epistemology. In particular, it can give us a new angle of inquiry on topics that currently seem stuck at an *impasse*. We saw in the previous chapter that neither conceptual analysis nor psychologism looks likely to offer progress with respect to intellectualism. Alongside this, there are two related topics I will consider in this chapter.

The first concerns *accessibility* – it is highly contentious whether we must, in some sense, have access to our knowledge. Philosophers who are broadly speaking internalists tend to endorse accessibility constraints, while those who are broadly speaking externalists tend to deny them.<sup>6</sup> It's worth noting that there are multiple notions of accessibility. A traditional idea put forward by 'language-first' philosophers is the articulation condition discussed in chapter 1: to know that p, one must be able to state that p when asked.<sup>7</sup> A second notion concerns *conscious access*: if one knows that p, one must consciously endorse the proposition.<sup>8</sup> A stronger constraint is that a subject must have conscious access to the *justification* for her knowledge, not just the content of what she knows.<sup>9</sup> Finally, some think of accessibility in terms of higher-order attitudes – that is, to know that p, one must know that one knows that p.<sup>10</sup>

---

<sup>6</sup> Of course there are many exceptions to this rule which is related to the fact there are many different notions of internalism and externalism. I am guided, though, by views like the internalism of Chalmers (2012) who whole heartedly endorses accessibility, and the externalism of Williamson (2000), who thoroughly rejects it.

<sup>7</sup> See Davidson (1975); Williams (1973)

<sup>8</sup> I referred to this as the introspection condition in chapter 1.

<sup>9</sup> See Smithies (2012); Bonjour (1998)

<sup>10</sup> This is the version of accessibility that Williamson (2000) attacks primarily.

Further, I will discuss the possibility of a safety constraint on knowledge – this is the idea that if one knows that *p*, one could not easily have been wrong about *p*. A strong version of this claim, defended by Williamson (2000), requires that knowledge states have a ‘margin for error’, entailing that the proposition known is not false in nearby (relative to an appropriate measure) possible worlds – and this is taken to be incompatible with many forms of accessibility. Other philosophers, such as Berker (2008), have argued for a weaker version of safety that is compatible with accessibility.<sup>11</sup> Others still, e.g. Neta and Rohrbaugh (2004), are sceptical of safety principles in general.

The verdict on these three topics has profound implications across epistemology (and philosophy more generally) and in all three cases, I think, conceptual analysis has stalled. As I discussed in chapter 1, there are internally coherent internalist and reliabilist theories of knowledge – which offer opposing verdicts on these topics. Conceptual analysis does not provide the means to decide between these two perspectives given that their proponents disagree as to what the correct intuitions are.

The practical role of knowledge attribution gives us an alternative means to decide between competing views on these topics. If one way of classifying the cases fits better with the practical role, this is good reason to take it to be correct. We will see below that different accounts of the practical role deliver different verdicts on these controversial topics – therefore, assessing which is correct is of pressing interest.

Before moving on, I also want to mention a couple of qualifications about how I am *not* understanding the project. First of all, I am not attempting to link this practical role to any genealogy of the concept of knowledge – imaginary or otherwise. This is in contrast to Edward Craig, pragmatism’s most influential proponent, who sketches a process whereby people living in a ‘state of nature’, initially without a concept of knowledge, would come to acquire one. Though fascinating, I think this approach raises a host of additional issues – is this imagined history purely heuristic? If it has genuine explanatory force does this derive simply from how things occurred in the story or is it meant to latch on to something about the actual history of our concept? I wish to put these issues to one side and focus on the role our concept plays *now*, no matter how we got here.

---

<sup>11</sup> Further, Srinivasan (2015) argues that such weaker principles are unsatisfactory.

A second point is that I don't want to try to do too much with the practical role. In particular, I don't want to try to provide a rationale for *all* intuitions about the application of the concept of knowledge. I will stick to places where the practical role detailed in broad strokes delivers clear results. A worry, if one tries to account for all the elaborate thought experiments – as Craig does – is that one slips into telling 'just-so stories'. One worries that similar stories could have been told if our intuitions had been any which way – and this deprives appeal to practical role of its explanatory force.

With this background out of the way, I want to look at what the practical role of knowledge is. A good place to begin is with the most thorough elaboration of pragmatism to date: Craig's *testifier account* (my term, not his). It is certainly rich enough to meet my adequacy conditions C1-C3. And though I think it's mistaken, it is instructively so.

### 3. Testifier Accounts

In *Knowledge and the State of Nature*, Edward Craig presents his account of the practical role of knowledge. He claims that the role of knowledge attribution is to identify 'good informants' – that is, people who can provide us with useful information. When we say someone is knowledgeable, we communicate that they are a good source of information.

The key issue, in elaborating such an account, is spelling out what constitutes a 'good informant'. The most straightforward interpretation of being an 'informant' is being someone who provides *testimony*, and the most straightforward reading of an informant being 'good' is being *trustworthy*.<sup>12</sup> This gives us the following:

*The Testifier Account:* The practical role of knowledge attribution is to pick out trustworthy testifiers.

I think this account has at least *prima facie* plausibility – additionally, it has been endorsed by a wide range of philosophers, which suggests it merits investigation.<sup>13</sup> Craig argues that identifying good testifiers is both useful, and a good fit with our actual

---

<sup>12</sup> I will get to more nuanced accounts below.

<sup>13</sup> All the philosophers mentioned in note 2 are sympathetic to Craig's account.

practice of knowledge ascription. The first point seems *prima facie* plausible. We all have an interest in acquiring true beliefs about various issues, and knowing who to ask to find out will help a great deal. Sometimes we need to know about a particular matter and are not well situated to investigate it directly – in such a case acquiring someone else’s testimony is invaluable. Consider the following example:

*Tribe Testimony:* Alice is a tribe leader and is worried about attack from a rival. She hears conflicting reports about her rival’s plans from passing merchants. One tells her the rival is preparing to attack imminently; another tells her the rival is waiting until spring. If her rival is attacking now, Alice must prepare her people to fight; if he is attacking in spring, she must spend the coming months building up the village’s defences.

Knowing which of the merchants is providing good testimony would be invaluable. She would then be able to base her response on accurate information. Thus, having a means to communicate when someone is providing good testimony is useful.

The key issue is whether such a practical role fits with our concept of knowledge – do we apply the term in such a way that it tends to pick out good testifiers? Being a good testifier certainly requires possessing a true belief, but one might worry that this is *all* that it requires – after all, Alice only cares about which merchant is telling the truth, not whether their belief is appropriately justified.

However, on reflection, it can be seen that this worry is misguided. The reason for this is that for testimony to be good, it must be *trustworthy* as well as true – we have to believe what we are being told in order to employ the information. If the merchant has a true belief on the basis of a lucky guess, Alice will have no reason to trust him. If she asked him what he based his claims on, all he could tell her was that he took a guess. She would not then adjust her plan and so would not benefit from his testimony. If, on the other hand, the merchant had overheard her rival talking with his soldiers, Alice could be confident the merchant’s belief was reliable and so trust it. Thus, we have an interest, specifically, in identifying people with testimony which we are able to identify as

reliable, in order to be able to act on it. This shows the *testifier account* fits roughly with our patterns of knowledge ascription – thus meeting C1.

A second point is that we don't just have an interest in identifying good testifiers, but in *promoting* them. The more of our peers that can provide good testimony, the better. This makes sense of the normative aspect of knowledge – since good testifiers are *praiseworthy*. This shows the testifier account meets C2.<sup>14</sup>

It's also worth noting that, if true, the testifier account would have some significant consequences – as we can see by relating it to the topics mentioned above. The first thing to note is that the testifier account suggests an *articulation constraint* on knowledge – for someone to know that p, they must be able to say that p if asked. For a state to be the basis of good testimony, the subject must be able to testify. As we've discussed, the articulation constraint is in itself a form of accessibility.

Moving on to the other kinds of accessibility: since the states we can articulate tend to be those with contents we have conscious access to, the testifier account motivates this constraint too. Finally consider the access to *justification* constraint. This fits well with the idea that good testifiers must be people we are able to trust. For us to trust someone's testimony, they must be able to provide us with the *reasons* they know it to hold. Being able to articulate the reasons for one's knowledge very much seems to involve having access to justification.

Moving on, the testifier account is in clear tension with intellectualism about know-how. There is much that we know how to do but are unable to say how we do it. Thus, a large portion of know-how cannot serve as a good basis for testimony. That someone knows how to do something does not necessarily make them a good testifier on the matter.

Finally turning to safety, I don't think the matter is so clear-cut. There does not seem to be a direct link between safety conditions and the testifier account. I think stretching the understanding of a good testifier in such a way that it provides a verdict on safety would be the kind of over-extension I vowed to avoid. Therefore, I'll take the testifier

---

<sup>14</sup> This point is emphasized by Dogramaci (2012).

account to be neutral on the matter – though I would be open to persuasion either way if a convincing rationale was offered.<sup>15</sup>

It would be significant for epistemology if the testifier account were correct, in virtue of its verdicts on accessibility and intellectualism. I think, though, that it is mistaken – it does not adequately capture the practical role of knowledge attribution. I will now consider some objections that attempt to show this. The first, I think, is unsuccessful – I consider it because it is a natural objection to all pragmatic accounts of knowledge, so seeing why it fails is important. The second, however, has teeth and ultimately pushes us towards my preferred version of pragmatism.

### 3.1 The Objection from Outliers

A natural objection to the testifier account is that there seem to be many cases of knowledge that do not play the necessary practical role. We ascribe knowledge to people who are currently incapable of providing testimony – e.g. because of paralysis. We also ascribe knowledge to people who are remote in space and time and so could never talk to us, and to our enemies who would never wittingly tell us the truth. Thus, it seems like the testifier account does not fit with our application of knowledge after all.

The appropriate line of response to this objection is outlined by Craig. He points out that the best form of pragmatism is going to identify knowledge with the state that *tends* to play the relevant practical role across a range of circumstances.<sup>16</sup> Though there are times when we need to find out whether a particular person can provide a particular piece of testimony in particular circumstances, our aims are generally more *amorphous*. It's useful to know, therefore, when people are *general purpose* testifiers, able to inform us across a range of circumstances. This suggests that the practical role outlined by the testifier account will be best served if knowledge ascriptions pick out states that tend to produce good testimony *in the absence of interfering factors*.

---

<sup>15</sup> One possibility would be to argue that the kinds of accessibility required by the testifier account are incompatible with some kinds of safety.

<sup>16</sup> See his (1990) chapter 10. He uses the term 'objectivization' to refer to the process of moving from a concept that solves particular goals in particular situations to a more generally and flexibly useful concept.

This idea can be developed further by looking at the normative angle. As Dogramaci (2012) argues, we can improve the testimonial practices of our peers by criticizing third parties. When we point out the knowledge (or lack of it) in people remote from us, we generally promote epistemic practices that lead to good testimony. Thus there is practical benefit to ascribing knowledge to people even when we stand no chance of benefiting from their testimony.

This shows, I think, that the objection from outliers is not convincing. Moreover, this general line of response is applicable for other versions of pragmatism.

### 3.2 The Limits of Testimony

A more compelling objection targets Craig's key thesis: that the practical role of knowledge ascription is primarily based upon testimony. I'll provide two reasons why the assumption is implausible. First, because transmission of information is not such a cut and dried notion as the initial outline of the testifier account might suggest, and doesn't apply only to states that can be articulated. Second, because there are many other features of representational states that are just as worthy of attention as testimony (and even transmission of information more generally). Testifier accounts miss the bigger picture.

It's clear that there are plenty of cases where someone possesses a bit of information they can't articulate and yet they are able to transmit it. One way of doing this is by *demonstration*. I may not be able to state how to swim breaststroke properly (beyond the very basics of the maneuverer) but I can teach a suitably receptive student by swimming correctly in front of them. Similarly, a jazz musician without formal training might not be able to explain verbally where a novice is going wrong, but could play the piece correctly to communicate the information.

As well as demonstration, one can transmit information by *guiding*. A dance instructor will correct poses by handling the student and rearranging her limbs appropriately; an expert carpenter might physically guide the apprentice's hand as she operates a tool. There is also non-verbal communication such as gestures and diagrams, used to get across information one is unable to put into words. Finally a teacher may use verbal *cues* that do not explicitly express the information she is communicating. For

example, a music instructor may not be able to state in full how to correctly play *Ornithology* but she can offer corrections and hints in response to various attempts by her student. In a lot of cases, information will be communicated by a mixture of all of these methods. This isn't just a matter of the teacher's abilities but of the *student's* too: it's well known that different people learn better by different means.<sup>17</sup>

### 3.3 Actionable Information

This suggests that if the practical role of knowledge concerns identifying people who can pass on information, focusing on testimony would be unduly restrictive. It would seem better to identify good informants with those who can 'provide information' in some more general sense. This suggests the following account:

*The Information Account:* The role of knowledge attribution is to pick out sources of actionable information.

The phrase 'actionable information' implies that the knower must be able to provide information that others are able to make use of themselves. Such a view is defended by Henderson (2009), and might be thought to be the most charitable reading of Craig himself.

This view has problems of its own, however. First of all, it is unclear, on this view, how to distinguish informants proper from people information can be derived from. For example, if you are sunburnt, I can use you as evidence to acquire the information that it was recently sunny. However, you did not *inform* me of this – and we would not ascribe you knowledge on this basis. With the testifier account, it was clear which information provided was relevant since it was simply that which arose from testimony. With the information account it's less clear. This also makes it harder to draw implications about controversial topics in epistemology, on the basis of this account.

---

<sup>17</sup> See Mayer & Masser (2003) on verbal and visual learners.

A more serious problem, though, is that there is more to the practical role of knowledge ascription than communication of information *by any means*. This will be the topic of the next section.

#### 4. The Significance of Action

The fundamental problem with both the testifier and information accounts is that the practical role of knowledge attribution goes far beyond identifying informants. Knowledge states affect a huge range of behaviour, and testimony is only a small part of this. I will show that we have a practical interest in a great deal of this behaviour, and our account of the practical role of knowledge ascription should reflect this. My argument proceeds in two stages: first I will show that we have a practical interest in our peers possessing a state that lines up roughly with our concept of knowledge; second I will show how this informs the practical role of knowledge *attribution*.

To put the first point simply: what people know determines how they attempt to achieve their ends. Moreover, if they have knowledge as opposed to mere belief, they are more likely to be successful – and this is a matter of great practical interest. Consider again the example of the tribe leader:

*Tribe Action:* The members of Alice's tribe are divided into hunters and gatherers. She has an interest in making sure they are successful at hunting and gathering respectively – if they are not, everyone will starve. In order to be successful, hunters and gatherers must possess certain bodies of information – e.g., where prey is located at various times of year, and what to wear to go undetected. This means that Alice has an interest in her tribe-members possessing the relevant information.

Note that since Alice has no need to go hunting herself, she has no interest in learning these things and thus no interest in whether the hunters are good *testifiers* – therefore, our practical interest in knowledge goes beyond its ability to yield reliable testimony, or reliable informants more generally.

This point generalizes. In a cooperative society like our own, it is for the most part in everyone's interest that everyone else achieves their ends. We want the engineers, pharmacists, plumbers, teachers, bin men etc. to achieve particular goals, so that our planes, medicine, sinks etc. work correctly. Therefore, we want them to have reliable information about the relevant subject matters. Moreover, we do not have any particular interest (apart from curiosity) in acquiring this information ourselves.

This establishes that it's in our interest for others to *possess* reliable information. Why, though, do we have an interest in *attributing* knowledge? The reason is that doing so allows us to predict success, and to intervene when failure looms. Suppose Alice learns that a particular gatherer is exceptionally knowledgeable about hunting. She can reassign him to hunting duty to maximize productivity. Alternatively, if she learns that a certain gatherer lacks the requisite knowledge about gathering, she can relieve him of his duties or arrange for him to be given the necessary training. Again this extends to a society like our own. We have an interest in knowing when people will be successful in their respective vocations and learning when to intervene and correct them if they're on track for failure.

This suggests that the practical role of knowledge attribution is linked to the identification of successful agents generally, rather than just agents who can provide us with information – as the previous accounts claimed. I'll develop this proposal in the remaining sections.

#### **4.1 The Action Account**

The previous section aimed to show that we have a practical interest in picking out states that are conducive to successful action; and to make the *prima facie* case that knowledge attribution fulfils such a function for us. In this section, I'll spell out the pragmatist account these claims motivate. The central idea can be summed up as follows:

*The Action Account:* The practical role of knowledge ascription is to pick out a state which tends to guide agents' action in a successful manner.

My task is to show that this gives us an account of knowledge meeting constraints C1-C3 outlined above: that we distinguish the role of knowledge from that of belief; that we explain the normative component of knowledge; and that we are able to use the account to resolve some significant epistemological stalemates. In meeting this challenge I have two main tasks. First I will show how the action account motivates identifying knowledge with a particular kind of state: robustly true information states that guide our attempts to realize our desires and that are sufficiently sophisticated. Second I will look at how such an account affects the epistemological disputes I have highlighted.

To begin, we must look at the *way* in which knowledge leads to successful action. Being tall leads to success in basketball, but tallness isn't a kind of knowledge. However, we saw earlier that the attitudes in general guide behaviour in a *systematic* manner – in particular, the functional role of an attitude state is systematically related to its content. Beliefs guide action in a way that would tend to lead to success if their content were true. This only applies, of course, when the state of affairs the belief is about is relevant to the particular task – that is, for a belief that *p*, success in the task at hand must depend on whether *p* holds.

Since knowledge entails true belief, this means that if *S* knows that *p*, *S* will tend to be successful in tasks where the outcome depends on whether *p* (I will call this the 'content-success relation'). For example, if I want water and I know that the nearest source of water is in the kitchen I will go the kitchen. Since the question of whether the nearest water is in the kitchen is relevant to the task of getting water, my knowledge will guide me successfully in this case. This gives us a systematic relation between knowledge and success.

So the action account requires looking for a state which satisfies the content-success relation. I will show that it is beliefs that are, in some sense, *robustly* true that meet this constraint. I take robustness to be a flexible notion that requires a kind of reliability (for now I'll leave it open whether this entails safety, sensitivity, or the like – I'm using the term so as to be neutral between the various accounts in the reliabilist ball park). If this association between robustness and knowledge ascription holds, we will have an account which meets C1, since it captures roughly our usage of 'knows', clearly distinguishing it from the role of belief ascription.

To understand the impact of the robustness condition, it's helpful to consider an argument that all that's needed for successful action is beliefs that are actually true. We are only interested in actual success, the argument goes, and all this requires is actual true belief, not true belief in various alternative scenarios. If this were in fact the case, the action account would be a poor fit with knowledge attribution – since we are unwilling to attribute knowledge in many cases of true belief.

However, the objection is mistaken. The reason for this is that for a belief state to guide an agent to successful completion of a task, it must be maintained and updated appropriately in the face of incoming evidence. To see this, suppose one of Alice's hunters has a true belief as to the location of some prey. For this to be useful, she will have to retain this belief until she catches the animal, rather than mistakenly changing her mind over the course of the hunt. This point is emphasized by Williamson (2006), who explains how successful action does not generally result from a single decision in which one's premises were correct, but through *continuous*, appropriate action guidance, with modification of one's plan where necessary.

We can see that many of the true beliefs that we do not take to be knowledge will be highly unstable, and insensitive to pertinent changes in circumstance.<sup>18</sup> True beliefs with no sort of legitimate basis – beliefs formed by a lucky guess, wishful thinking etc. – will clearly be ill-suited to guide action over an extended period of time. There is no reason for these beliefs to be updated as required in the face of incoming evidence. If one keeps on guessing, one will run out of luck; with wishful thinking, one will likely find a situation in which the way the world is and the way one wants it to be do not match.

The same holds for many of the paradigmatic cases of justified true belief we do not take to be knowledge. Take a canonical Gettier case: Jones' true belief that either Smith owns a Ford or Brown is in Barcelona.<sup>19</sup> This belief is unstable, since it is founded on the false belief that Smith owns a Ford – if Smith were to mention he was driving a rental (which he might well do) Jones would abandon his belief. Moreover, if circumstances were to change so that the proposition became false (e.g., because Brown returned to work), Jones would not revise his belief accordingly. Thus, Jones' belief cannot be relied

---

<sup>18</sup> This point is emphasised by Williamson (2000) – I take it to be, in part, the moral of the diamond thief example.

<sup>19</sup> This is found in the original paper by Gettier (1963).

upon to successfully guide action. This shows that the action account of knowledge attribution requires it to pick out a state that meets a robustness condition.

Another objection is that it seems like there can be knowledge states that do not play any role in guiding successful action – for example, if the knower is paralyzed and so cannot act at all, or if the knowledge concerns some completely useless subject matter – which seems inconsistent with the action account. However, on reflection, it's clear that the action account *can* allow for such cases. As we discussed above, the best form of pragmatism is going to identify knowledge with the state that *tends* to play the relevant practical role across a range of circumstances.

Robustly true information states tend to produce successful action – provided, of course, a situation arises in which the success of an agent's action depends on the truth-value of the relevant proposition. We have a general interest in attributing such success-conducive states in a systematic manner without worrying about whether a particular kind of success is attained on a particular occasion. This follows from the fact that we need to systematically communicate information that will be useful across a range of eventualities, as well as helping out with specific goals in specific circumstances.

#### **4.1.1 The Force of the Argument**

Before moving on, I want to consider how much *force* my argument for the action account can be assigned. One might worry as to the extent to which pragmatism really provides fresh *traction* on the investigation of knowledge. It might be thought that my defence of the action account begs the questions that stalled conceptual analysis. That those with intuitions that prioritize the role of assertion and conscious judgment will see identifying informants as the most important aspect of the practical role of knowledge attribution; while those with intuitions that prioritize action will focus on the identification of successful agents. Thus, the deadlock remains.

This argument is based on a misconception: pragmatism does not look to what account of the practical role of knowledge attribution is intuitively attractive but to what the practical role *actually is*. As Block and Stalnaker pointed out, this an empirical matter determined by the role our use of language plays in a social setting. Importantly, we can

see that pragmatism brings a new kind of evidence to bear on the debate: the relationship between our language use and our social structure.

We can imagine different societies, which favour the different accounts. On the one hand, consider a libertarian utopia in which rugged individuals all live entirely off the land they own – they do not rely on the work of others to survive. They do, however, regularly convene to discuss tips on how to get the most from their land. Presuming these people make knowledge attributions during their conventions, their main practical concern will be identifying good informants.

On the other hand, consider a cooperative society, in which there is division of labour, and collective ownership and consumption of all the society's products – so that everyone requires mutual success to survive. However, in this society, no one is willing to ask for help or take advice on how to achieve their goals – consider it the logical extreme of proud refusal to ask for directions. In this case, people would have a practical interest in identifying successful agents, but not good informants.

If our society were like the first, the testifier account would be correct, while if it were like the second, the action account would be. The facts about our society are less clear-cut since we have a practical interest in both testimony and successful action – however, I think they're still clear enough to vindicate the action account. In almost every aspect of our lives, we rely on the successful action of our peers in one way or another; however, we require testimony only sporadically. This holds even when we restrict our attention to states that satisfy the articulation constraint. As was noted in chapter 1, these states *also* guide action, and so their practical significance extends beyond the cases in which they are articulated to their role in producing successful action: I could provide testimony as to the location of Morpeth, thanks to my knowledge that Morpeth is in the north of England; however this knowledge also allows me to navigate successfully, which is of wider practical significance.

Further, the fact that knowledge attributions identify good informants is a *corollary* of the fact that knowledge attributions identify successful agents. The reason for this is that one task an agent will often have is to help her peers succeed, and providing reliable testimony is a way of doing this. Thus providing testimony is one element of being a successful agent in general. Because of this, though Craig and his followers have latched

onto something importantly right about the role of knowledge attribution, they're missing the more fundamental picture. Crucially, I think these observations about practical concerns in our society do not beg any questions against the philosopher with articulation-first intuitions.

## 4.2 Sophistication

I have argued that the action account requires an information state to be robustly true to count as knowledge. A natural question is whether this is *all* that's required. It's plausible to think that some such information states are more practically significant than others. A claim familiar from debates on where to draw the distinction between belief and subdoxastic states is that beliefs are more *sophisticated* (in some significant kind of way) than 'mere information'. This suggests that the action account should impose additional constraints on knowledge.

The kind of sophistication condition I have in mind aims to pick out states that are systematic in their functional role, in virtue of being in some sense *integrated* with other information. There are a number of ways to try to specify the relevant type of integration: one aspect, emphasised by Gendler (2008a+b), is evidence sensitivity; another is inferential integration, emphasised by Stich (1979) and Levy (2015); a related idea is the generality constraint, advocated by Evans (1982) and Davies (1989). These conditions ensure that the states in question will guide action in a coherent manner, generally responsive to the state of the environment. This is practically significant for two key reasons: first, knowing about the presence of such states in subjects allows us to predict success in a systematic manner; second, we can *influence* such states systematically. Also note that the states that fail to meet such a condition – for example, those involved in pre-perceptual processing – are not things we naturally describe as knowledge.

That we have particular interest in states with systematic action guiding roles should be clear. It means we can predict a wide range of action on the basis of limited information. It also means we can predict how a subject's action will change in response to new environments. The significance of their susceptibility to influence stems from the fact we have an interest in *promoting* success, not just predicting it – as was mentioned

above. If a subject possesses information states that are evidence sensitive and inferentially integrated but are not currently robustly true, one can *make* them reliably true by presenting appropriate evidence. That such states are of particular significance follows from the cardinal rule of pragmatism: focus on the things you can do something about. There is more to be said about this notion of sophistication, but I will postpone looking at it in more detail until the next chapter.

This leads us, I think, to an attractive account of knowledge. The robustly true and sufficiently sophisticated information states which I wish to identify with knowledge are the states that lead to successful action in a systematic manner. They are therefore states we have an interest in both knowing about and promoting. Moreover, the fact such states are open to influence makes sense of the normative component of knowledge. Criticizing and praising with respect to such states can lead to *improvement*. This shows the action account meets C2; I'll now look at the third and final constraint.

## **5. Consequences**

If we accept that the action account gives the correct account of the practical role of knowledge attribution, what does this tell us about the nature of knowledge? In this section, I will look the epistemological topics I've highlighted, and see what implications the action account has for each of them. Importantly, the action account delivers different verdicts from the testifier account in all cases. This shows the significance of the pragmatist project for epistemology generally. Even if you are not convinced by my arguments in favour of the action account, you should be invested in whether it or the testifier account is correct if you have an interest in the topics listed below.

### **5.1 Accessibility**

The first question is whether there is an accessibility constraint on knowledge. The action account says that 'knowledge' picks out states that tend to guide successful action in a systematic manner. Thus, our question is whether such states must satisfy an accessibility condition. As was discussed above, there are multiple notions of accessibility. The first is

an ability to articulate the knowledge. The second is conscious access to the information. The third is conscious access to one's *justification*, so that one can provide reasons for endorsing the proposition.

I think the action account provides good reason to reject all these versions of the accessibility constraint. We have successful action guiding states that we have no conscious access to and cannot provide reasons in support of – these are the states that guide *skilled action*. One might object that such states, though success-conducive, are not sufficiently sophisticated to meet the criteria of the action account. Though this is a serious challenge, I think it is one that can be met. The argument for this comes in the next chapter, when I discuss the different criteria for sophistication.

The final notion of accessibility is a little trickier, since a lot turns on what higher-order knowledge amounts to. One might think it requires conscious accessibility, and so is to be rejected on the basis of the current considerations. However, it is held by some theorists that general facts about representation entail that any knowledge state trivially comes with higher-order knowledge – on this see Greco (2014). Therefore, I think we should set this issue to one side.

The important point to take away is that the action account is in tension with the internalist accessibility constraints that the testifier account favours.

## 5.2 Intellectualism

When looking at how the action account relates to intellectualism, similar considerations arise. Paradigmatic cases of know-how involve sophisticated states that provide success-conducive action guidance. The success conduciveness is clear: if I aim to get to work on a bicycle, knowing how to ride a bike will help me to do so successfully.

I also think we can show that such know-how is appropriately sophisticated. First, we can establish that know-how interacts with desires to some extent. This follows from Railton's (2009) discussion of how various skilled actions display what he calls *fluent agency*, so that our performance of them is sensitive to our current mood and aims. For example, if I'm cycling and running late I can ride quickly; if I'm angry I can cycle aggressively; and if I want to impress my friends I can incorporate daring stunts.

Moreover, these skilled actions are learned in part through perceptual feedback gained in training – and this seems like a kind of evidence sensitivity. They also display more immediate evidence sensitivity in adapting to new features of any particular circumstance: a bumpy road, high winds, unexpected obstacles etc. (Again the arguments in chapter 3 are needed to settle this conclusively.)

This shows that, according to the action account, there is know-that *involved* in typical cases of know-how. A couple of questions remain before it can be seen whether intellectualism is fully vindicated. First, there may be basic cases of know-how (such as knowing how to make a fist) that are not sufficiently sophisticated. Second it may be that know-how involves some non-sophisticated ability as well as the relevant know-that. I will not settle these issues here, since it would require an involved examination of know-how which would take us far afield from the current topic. We can see, however, that the action account leads to a view of know-how that significantly differs from the one suggested by the testifier account.

### 5.3 Safety

The safety constraint entails that for a belief to qualify as knowledge, it must be true not only in the actual world, but in a range of nearby possible worlds as well. The action account, I think, provides support for at least some version of the safety constraint. The most pertinent form of safety is *forward looking*. For a belief to be conducive to successful action, it must be able to continue to guide action appropriately until success has been achieved. If one is likely to revise a currently true belief in the near future, it does not meet this condition. As I explained above, many Gettier cases involve beliefs that lack forward looking safety – and this is why they do not amount to knowledge, according to the action account.<sup>20</sup>

However, there is also motivation for thinking that backward looking safety is vindicated by the action account – that is, true beliefs formed by an unsafe process are

---

<sup>20</sup> An important part of forward looking safety, emphasized in recent work by Das, is safety from future *defeat* – that is, that it is unlikely that new evidence will come along to make an agent mistakenly retract her belief. We saw above, that many Gettier cases lack this kind of safety. See Das (ms) for an argument that this is the principal safety condition on knowledge.

not knowledge. This is because, in line with Craig's notion of objectivization, the uses for knowledge attribution are general, concerning broad social goals. We use knowledge attribution to give credit to states which generally lead to successful action, and thus that should be encouraged. If a state was formed by an unsafe process, then imitating it will lead to failure, and thus it should not be praised. Further, since the action account does not motivate accessibility constraints, it is compatible with Williamson's strong construal of safety.

## **6. The Practical Role of Belief Attribution**

In this chapter, we have been discussing knowledge, but closely analogous ideas apply when it comes to the practical role of belief attribution. I chose to focus on knowledge because, thanks to Craig, that is what the most detailed prior theories concern. However, I think the motivation for pursuing a pragmatic account of belief is just as compelling. (I will go over this fairly quickly, since it should be clear how the detailed discussion of knowledge carries over to belief.) The question of where to locate belief within a detailed picture of our psychology is, again, not something either conceptual analysis or psychologism look well placed to answer. Moreover, belief attribution, like knowledge attribution, appears useful and so its practical role is worth investigating.

Philosophers have paid attention to the practical role of belief attribution. Most prominent is Dennett: as we discussed above, he claims that the purpose of belief attribution is to allow us to predict and explain behaviour systematically and efficiently. A number of other philosophers working on belief make statements broadly sympathetic to pragmatism. Schwitzgebel (2010, pp. 546-547) says 'The practical question is this: Do we want to highlight this empirical fact about ourselves – what I'd call the gulf between occurrent judgment and dispositional belief – or do we want to marginalize it as anomalous?' (and he takes this question to determine how the term 'belief' is to be applied). Smithies (2012, p. 352) says 'we need not concern ourselves with questions about our ordinary use of the term 'belief'... instead, we should ask which of these concepts are most useful for us'. Gendler (2012, p. 799) states '[I] make the simple claim

that it is conceptually useful to have a single category for the family of mental states that accompany these propensities’.

However, to the best of my knowledge, none of them go into the kind of detail for the practical role of belief attribution that Craig offers for knowledge.<sup>21</sup> We can see, though, that issues for the role of belief will arise, analogous to the ones we discuss with knowledge. Famously, Dennett’s ‘minimal’ account threatens to over-generalize in the case of beliefs too – it seems to justify ascribing beliefs to everything from our pre-perceptual modules to thermostats and alarm clocks. Further, beliefs too seem to have a normative component: not all beliefs are praiseworthy, but they are subject to criticism as *rational* and *irrational*.

Finally, there are philosophical topics we would hope the practical role of belief can offer insight on: the status of various accessibility conditions; where to draw the doxastic/subdoxastic boundary; and how to resolve the tension between articulation constraints and action constraints on belief. Thus we have criteria analogous to C1-C3 above.

In response to this, there is the potential for a *testifier account* and an *action account* of the practical role of belief. One could suggest that the role of belief attribution is to identify testifiers regardless of whether they are trustworthy or not. As Dogramaci (2012) has argued, it is in our mutual interest to improve the testimony of our peers – and this can be done by criticizing beliefs that fail to be knowledge as *irrational*. We need belief attribution to promote knowledge, since we need a way of identifying the things that are not yet knowledge.

The *action account* for belief can be motivated along the same lines as with knowledge. We are interested in knowing about the behaviour of others, not just their testimony. Knowledge attribution informs us about successful action, but we need belief attribution to know when and how people may fail in pursuing their plans. Further, we have an interest in intervening when people are looking likely to fail, in order to promote the cooperative goals – that is why beliefs that fail to be knowledge are apt for criticism.

---

<sup>21</sup> This is why, in chapter 1, emphasized the importance of being *explicit* about one’s commitments to *pragmatism*, to allow us to state upfront the constraints it puts on our theory (that is, C1-C3 in this chapter). I’ll show in the next two chapters how a systematic pursuit of pragmatism leads one to an importantly different account of belief than the authors who endorse but do not emphasize it so much arrive at.

Further, there is also motivation for an analogous sophistication condition – we have a particular interest in being able to predict behaviour systematically and efficiently. As we have mentioned, this sophistication also provides the possibility for systematic *influence*. Since we have a general interest in promoting successful action amongst our peers, we have a general interest in identifying the states that we can turn into success-conducive states in a systematic and efficient manner.<sup>22</sup>

## 7. Conclusion: Action First Attitudes

Now let's bring this all together. At this point we are in a good position to state the general *action account of the attitudes* – as motivated by pragmatism. The central planks of my view are in place. In the chapters that follow, I will mainly be filling in the details, and exploring the implications for a range of topics in philosophy.

1. *Knowledge*: The action account states that the practical role of knowledge ascription is to identify agents who will tend to be successful in particular ways.
  - From this we can conclude that knowledge states are robustly true sufficiently sophisticated action guiding states.
2. *Belief*: The practical role of belief ascription is to identify how agents will attempt to achieve their goals more generally, and inform us how we could intervene to lead them to be successful if necessary.
  - From this, we can conclude that belief states are sufficiently sophisticated action guiding states.<sup>23</sup>
3. *Rationality*: Finally, it's plausible to think that the role of judging people or belief 'rational' and 'irrational' is to influence people's beliefs to maximize knowledge and so promote successful action.
  - There's a lot more to say about rationality, but I will postpone talking about this until chapter 5.

---

<sup>22</sup> Since there are of course cases in which we either do not wish to correct someone's beliefs or are unable to, the same response to the objection from outliers I presented in case of knowledge is applicable here.

<sup>23</sup> An implicit constraint is that these action-guiding states must have the appropriate functional role for knowledge/belief, given their content, as was mentioned in section – I'll discuss this functional role in more detail in chapter 4.

I think this view is highly appealing, given the unified account it offers of the role of knowledge, belief and rationality. It is all based around our mutual interest in the successful action of our peers. Knowledge states are those that tend to lead to successful action, and thus are those we have a general interest in promoting – knowledge attribution allows us to pick out such states so we can promote them, and know when we can count on success. Beliefs are the states that would be knowledge if all went well – and they can systematically be turned into knowledge with the appropriate influence. Belief attribution allows us to pick out these states so that we know how people will attempt to fulfil their tasks, and they are the states we should correct if we want to promote successful action.<sup>24</sup> Finally, the concept of *rationality* is a tool we can use to ascribe praise and blame to beliefs, to make sure as many of them as possible end up being knowledge – which again promotes successful action.

Clearly, this account has many similarities with Williamson’s knowledge-first philosophy – he too endorses intellectualism, rejects accessibility constraints, and endorses safety. This is no surprise since Williamson assigns the relationship between knowledge and action a central role in his epistemology.<sup>25</sup> My account also lends itself towards seeing knowledge as prior to belief – it starts by looking at how knowledge promotes successful action, then identifies belief as the state we have an interest in turning into knowledge. Like Williamson too, I begin with *successful* action, rather than *attempts*.

One important contrast though, is that I first look at action and then work back to knowledge, whereas Williamson (2015) suggests we should do the reverse. Thus, my account can be thought of as *action-first*, as opposed to knowledge-first.

Williamson’s brand of broad externalism account is sometimes thought of as an unpleasant truth, forced upon us by proper theorizing – that it would be somehow preferable if Cartesian internalism could be made to work. However, my account shows why it’s something to be embraced: this concept of knowledge is an integral part of a cooperative society.

---

<sup>24</sup> Of course, we also have an interest in predicting when and how people will attempt and fail to achieve there ends, so that we can plan or react accordingly – even we lack either desire or ability to correct them.

<sup>25</sup> See in particular Williamson (2015).

To make an informed decision about my account though, we need to get clearer on the details of what it tell us about the mind. To do this, we need to better understand the nature of 'sophistication', and it is to this issue that I will now turn.

## Chapter 3

# Against Elitist Theories of Belief

### 1. Introduction

A recurring theme in the two previous chapters has been the tension between articulation based constraints on the attitudes, and action based constraints. I have argued for an action based account of the *practical role* of the attitudes, but this does not tell us how to ascribe attitudes in a range of cases. The reason for this is that the action account specifies that it is only behaviour that is sufficiently *sophisticated* that we have a general practical interest in. To get a more concrete sense of the nature of belief and knowledge on my preferred view, we have to see what kind of states meet this sophistication constraint.

I propose to look at this issue by considering cases of conflicting behaviour: scenarios in which a subject is disposed to assert one thing, and indeed endorse it upon reflection, while behaving as though she believes the opposite – for example, an elite sportsplayer who makes incorrect statements about how one should play. Therefore, there must either be some beliefs that are not manifested in verbal behaviour or there must be some action guiding states that are not beliefs. We can use the sophistication condition to evaluate what these subjects actually believe (it makes sense to look primarily at belief rather than knowledge here, to avoid worrying about the complication that one cannot know something false).

I will examine the argument that, in cases like these, only states involved in linguistic and conscious processes are *sophisticated* enough to qualify as beliefs. A number of philosophers have used this kind of argument to defend a restrictive account of belief that satisfies (something like) an articulation constraint – I call these *elitist* theories of belief. I will argue that elitism is mistaken. In short, the reason for this is that verbal states are

less sophisticated than might initially be supposed and non-verbal states are more so: both meet criteria for sophistication *imperfectly*. Therefore, both look to be equally good candidates to qualify as beliefs.

My argument will proceed as follows. First I will describe some cases of conflicting behaviour and the account of them offered by elitists (§2+3). Then I will look at two ways to make the argument for elitism precise, defended by Gendler (2008a) and Stich (1979) – who focus on evidence sensitivity and inferential integration respectively as criteria for sophistication (§4+5). My response will be to look at a number of examples that illustrate the imperfect ways in which our mental states meet these conditions. I will conclude by suggesting that we should move away from the elitist view entirely and adopt an *action* based account of belief.

## 2. Cases of Conflicting Behaviour

The examples of concern are ones in which a subject's verbal behaviour and conscious judgments indicates belief in a proposition, while much of her non-verbal behaviour indicates a belief in that proposition's negation. As I've mentioned before, conscious judgment is often grouped together with linguistic behaviour as the internal analogue of assertion – a subject will consciously affirm a proposition to herself when she's unwilling or unable to assert it aloud. I'll call verbal behaviour and conscious judgment *elite behaviour*, and other kinds *non-elite behaviour*.<sup>1</sup> Now, consider the following cases:

*Skywalk*: Over the Grand Canyon there is a giant glass bridge that members of the public may walk out onto. The average subject will be well informed of the strength of glass of this thickness; moreover she will sincerely assert the sentence 'I am in no danger, the bridge will support my weight'. And yet despite this, while on the bridge she trembles and sweats and is eager to get across it as soon as possible.<sup>2</sup>

---

<sup>1</sup> It would be more fitting to use the terms 'intellectual behaviour' and 'non-intellectual behaviour'. Unfortunately, as we've seen, 'intellectualism' has been established as the thesis that know how is know that. This makes what I'm calling 'elitism' closely analogous with *anti-intellectualism*. It's necessary to use this conative term to avoid unnecessary confusion.

<sup>2</sup> This example comes from Gendler (2008a).

*Implicit Bias*: Psychological studies reveal that many white Americans who profess to being committed to racial equality discriminate against black people in their unconscious behaviour. They sincerely assert that black people deserve equal treatment, are equally trustworthy etc. And yet they are less willing to make eye contact, stand further away and display other ‘micro-aggressions’ during social interactions with black people; they are also less likely to hire black candidates relative to the quality of their CV and discriminate in numerous other ways.<sup>3</sup>

*Roadworks*: Ben is told that due to road works, the bridge he normally takes on his way to work will be closed. Upon hearing this, he thinks to himself (consciously) that he will have to take the roundabout route. He is disposed to sincerely assert this to anyone who asks him how he’ll be travelling to work in the next week. However, when he drives to work, he’s disposed to set off on the old route, not leave extra time, etc.<sup>4</sup>

The subject on the skywalk has elite behaviour suggesting she believes that she is safe, while her non-elite behaviour (her trembling and sweating) suggests a belief that she’s in danger. The implicitly biased subject’s elite behaviour indicates belief that black people are equally trustworthy and deserve equal treatment to white people, while the unconscious behaviour indicates a belief that black people are worthy of suspicion. Ben’s elite behaviour indicates a belief that the bridge is closed while his non-elite behaviour indicates a belief that it is open. The question, in light of this is: what do the subjects actually *believe*?

To put things neutrally, the subjects have two ‘belief-like states’ – the one guiding elite behaviour and the other guiding non-elite behaviour. I’ll refer to these as *elite states* and *non-elite states* respectively. So our question can be rephrased as asking: which of these belief-like states are actually beliefs? Both the elite and non-elite states have an

---

<sup>3</sup> See, e.g., Gendler (2008a), Schwitzgebel (2010), Gertler (2011) for discussion.

<sup>4</sup> This example comes from Schwitzgebel (2010). Gendler’s (2008a) case of the lost wallet, and Zimmerman’s (2007) case of Hope and the Dustbin are similar in form.

action guiding role appropriate for belief, while (by construction) only the elite states have contents the subjects can articulate.<sup>5</sup>

### 3. Introducing Elitism

Elitism is the thesis that the subjects believe only what they say (and consciously affirm) in these examples – or more generally the claim that only elite states are beliefs.<sup>6</sup> So according to the view, the subject on the skywalk believes that she's safe, the implicitly biased subject believes that all races are equal, Ben believes that the bridge is closed – and none of the subjects believe the proposition indicated by their non-elite behaviour. Though there are many dissenters,<sup>7</sup> elitism has been influentially advocated by a number of philosophers. Here are some notable endorsements of the view:

[I]f a subject is psychologically (and physiologically) normal, inclined to be cooperative and has no motivation to deceive us, then if she believes that *p* and is asked whether *p* is the case, she will generally say that it is... [In some cases] the subject may be temporarily paralyzed and thus unable to assent to anything. Or he may have a strong desire to mislead his questioner, or simply wish to say nothing. Still, under these circumstances, if we ask a subject whether *p* is the case, he will generally have a certain sort of characteristic experience... One might also describe the experience as being aware that *p* or being conscious that *p*. [Stich (1978) pp 40-41]

[In cases of conflicting behaviour] our beliefs and desires mandate pursuing behaviour B and abstaining from behaviour A, but we nonetheless find ourselves acting – or feeling a propensity to act – in A-like ways... It seems misleading to

---

<sup>5</sup> I'm assuming that assertion is itself a type of action.

<sup>6</sup> My arguments apply equally to the broad and narrow version of the thesis so I will not dwell on which is the more plausible.

<sup>7</sup> There are many ways one could make such a denial. For example: Shoemaker (2009) claims that both elite and non-elite states are beliefs, so that the subjects in the examples above have contradictory beliefs; Schwitzgebel (2010) argues that elite and non-elite states are both parts of a single complex belief state, so that subjects with conflicting behaviour are in a state of 'in between belief'; Gertler (2009) claims that the non-elite states alone are beliefs.

describe these A-like behaviours as fully intentional [i.e. as manifestations of belief]: their pursuit runs contrary to what our reflective commitments mandate. [Gendler (2012) p 799.]<sup>8</sup>

[B]eliefs and other cognitive states are analyzed in terms of their dispositions to cause phenomenally conscious episodes of judgment, rather than their dispositions to cause physical behavior. [Smithies (2012) p 348]

[T]he most straightforward expression of a belief is an assertion...beliefs, recognitions and so on, are going to be ascribed to animals in an impoverished and ... a somewhat conventionalised sense. [Williams pp 139-140]<sup>9</sup>

Granting the action account of the practical role of belief ascription, the most promising line of argument for elitism appeals to the notion of sophistication. The action account entails that belief requires a certain degree of sophistication. If one can show that only elite states meet this condition, one will have a compelling argument for elitism.

Sophistication is a loose notion, so one needs to work with a more precise criterion for belief if one is to assess this line of argument. Two promising strategies here have been to appeal to evidence sensitivity and inferential integration as necessary conditions for belief.<sup>10</sup> These seem like a reasonable gloss on sophistication, since responsiveness to the environment and to previously obtained information distinguish an intelligent process from an automatic one. It also seems *prima facie* plausible that elite behaviour has a privileged connection with these features. For example, it is through reading and writing that scientific opinions are expressed and communicated. This linguistic behaviour is sensitive to incredibly complex experimental evidence, as well as the detailed testimony

---

<sup>8</sup> Gendler claims that the non-elite states I'm considering along with many others form a distinctive mental *kind* she calls 'alief'. It should be noted that my arguments do not entail that *all* cases Gendler classifies as alief are beliefs since many of them – such as priming effects – lack the appropriate degree of sophistication or action guiding role. Thus I leave room for a less expansive notion of alief.

<sup>9</sup> See also Davidson (1975), Brownstein & Madva (2012) and Zimmerman (2009) for similar ideas.

<sup>10</sup> There have been other attempts to capture the complexity of belief similar in spirit to the ones I'm considering – for example, Davies' (1989) appeal to the generality constraint. I think the objections I raise will carry over to alternative formulations. See note 30 below for further discussion.

of other researchers; and, it is in writing that researchers tend to work through long chains of inferences when, for example, doing mathematical proofs.

Despite their plausibility, I think both versions of this argument fail, as I'll now argue. Both elite and non-elite states are *imperfectly* evidence sensitive and inferentially integrated. Thus, considerations of sophistication do not count in favour of elitism.

#### 4. Evidence Sensitivity

The argument centred on evidence sensitivity is given by Gendler, who writes:

[W]hatever belief is—it is normatively governed by the following constraint: belief aims to ‘track truth’ in the sense that belief is subject to immediate revision in the face of changes in our all-things-considered evidence... In each of the cases we have been considering, only one of the competing tendencies is evidence-sensitive in this way. The man [on the skywalk] believes that he is safe because if he were to gain evidence to the contrary, his attitude would change accordingly.<sup>11</sup>

To state the argument explicitly:

- 1) All beliefs must be appropriately evidence sensitive
- 2) Only elite states are appropriately evidence sensitive
- C) All beliefs are elite states

For this argument to be successful, the elitist must specify what level of evidence sensitivity is ‘appropriate’ for belief. A promising starting point is the fact that in all the cases of conflicting behaviour we’ve discussed, only the elite state is sensitive to *linguistic* evidence. For example, the subject on the skywalk’s disposition to assert she’s safe is sensitive to whether the staff tell her that she is; while her trembling and sweating

---

<sup>11</sup> Gendler (2008b) pp. 565-566. In a précis of her work on belief (2012, p 763) she says: ‘beliefs are, roughly speaking, evidentially sensitive commitments to content that are quickly revisable in the face of novel information’ – so this appears to be her considered position. See also Brownstein and Madva (2012) p71 who make a similar argument concerning cases of implicit bias.

is unaffected by such testimony. One might argue, therefore, that beliefs must be sensitive to *all* the evidence, and since the non-elite states are not sensitive to linguistic evidence they cannot be a manifestation of belief.

Sensitivity to all evidence is a very stringent condition for belief so one might be tempted to weaken it. One option would be to go with the minimal constraint that beliefs must be sensitive to *some* evidence – and argue that non-elite states do not even satisfy this. Another option would be to adopt an intermediate condition: that beliefs must be sensitive to evidence in a sufficiently uniform and consistent manner; it might be thought that our ability to grasp linguistic evidence in a systematic fashion indicates this. Or, following Gendler, one could emphasise immediacy – she claims that belief is ‘subject to *immediate* revision’: that it can ‘turn on a dime’.<sup>12</sup> This is meant to contrast with the slow habitual changes in non-conscious behaviour. Here, then, are four alternative theses that could be used to flesh out the argument for elitism:

E1. Beliefs must be sensitive to all evidence.

E2. Beliefs must be sensitive to evidence in a unified and systematic manner.

E3. Beliefs must be immediately revisable in response to evidence.

E4. Beliefs must be sensitive to some evidence.

The elitist must then argue that the elite states meet the condition of choice while the non-elite states do not, thus securing premise two of the argument. None of these theses can be upheld however: even elite states fail to meet the criteria of E1/E2/E3, while many non-elite states satisfy E4.<sup>13</sup> I will first concentrate on the relationship between *linguistic* behaviour and evidence and then move onto other types of elite behaviour.

#### **4.1 Linguistic Behaviour**

Consider the following examples:

---

<sup>12</sup> Gendler (2008b) 565-566. She repeats this claim in her (2012) – see note 11.

<sup>13</sup> This line of thought is endorsed by Schwitzgebel (2010).

*Bob the baseball fielder:* Bob is waiting to catch the baseball. As the ball is struck he is unable to say where it is going to land; however, if a reliable source were to tell him the coordinates of where it was heading, he would then be disposed to relay this information when asked. On the other hand, the visual evidence of seeing the ball's trajectory after it's hit does not help; he is still unable to state the location – though he is able to move himself to the right place in order to catch it.<sup>14</sup>

This illustrates that often our elite states are sensitive to linguistic evidence but not to perceptual evidence. When it comes to the issue of where the ball is going to land, only Bob's non-elite states are sensitive to the relevant visual evidence.

*Sam the teenager:* Sam makes assertions about what work he must do to get a good job. He says things like “I don't need to work much at all to get a job: my friends' older brothers all did far less work than me and they're all lawyers and bankers now.” However, the elite states underlying these assertions are not sensitive to all linguistic evidence. If his friends were to tell him that the job market has changed and he needs to get some internships if he wants to succeed, he would revise what he said. However, if his family were to tell him the very same thing he would not change at all.<sup>15</sup>

I take it that this type of example is not unfamiliar. It illustrates that sensitivity to linguistic evidence is not an all or nothing matter even amongst elite states.<sup>16</sup>

---

<sup>14</sup> This example comes from Stalnaker (1991)

<sup>15</sup> Variations on this example are possible where what decides whether Sam is sensitive to linguistic evidence is not his social relation to the speaker but whether the speaker talks in his teenage vernacular, is sufficiently rhetorically adept, charismatic etc. This further illustrates the messiness of our relation to linguistic evidence.

<sup>16</sup> One might object that it could be that Sam's elite states are sensitive to the testimony of his parents; it's just that he has a standing belief that they are unreliable testifiers, which defeats such evidence. However, we can imagine that Sam is disposed to sincerely assert that his parents are reliable sources of information, and yet ignore their advice regardless – and the elitist is committed to saying that he believes what he says. I'm sure this is in fact the case with implicitly biased subjects – they say, e.g., that the testimony of black people is equally reliable and yet they completely ignore it.

*Rich the carnivore:* Rich asserts that there's nothing objectionable about eating meat, and indeed he eats lots of it. No amount of testimony as to how cruel the meat industry is will make him change what he says. However, if he were to visit a slaughterhouse and *see* the cruelty (the very same processes that he has already been told about) what he'd be disposed to assert would change.

This shows that what we are willing to say and endorse may not be sensitive to testimony but only perceptual evidence. Again, I take it to be a common phenomenon that seeing something will change our behaviour whereas being told about it will not.

*Thomas the slow learner:* Thomas is taking a set theory course and struggling a little bit. He's been told some strange things that he didn't believe before, such as that there are 'different sizes of infinity'; and he's been shown watertight proofs that demonstrate the claims. When he hears these things first time round he's just confused and doesn't believe what he's told. However, he is studious and reads through the proofs multiple times until eventually he 'gets it' and believes the theorems.

This shows that linguistic behaviour is not always *immediately* revisable in the face of evidence which mandates a change. Sometimes (especially in cases where the fact being indicated is especially radical, surprising or strange) it can take a while before what we've seen or been told *sinks in* and we finally change what we are disposed to assert. This phenomenon has been identified and studied in social psychology, and is known as *belief perseverance*.<sup>17</sup>

These examples together show that elite states fail to satisfy the criteria set by E1-E3 – at least when we restrict our attention to their manifestations in linguistic behaviour. Sam the teenager, Rich the carnivore and Bob the baseball player all illustrate how elite states are not sensitive to *all* evidence (E1). Moreover, Rich shows how elite states may not be sensitive to *linguistic* evidence, and Sam shows that our interactions with linguistic

---

<sup>17</sup> See Anderson (2007).

evidence may be messy (E2). Finally, Thomas shows how linguistic dispositions need not be *immediately* be sensitive to evidence (E3).

#### 4.2 Other Elite Behaviour

A natural response to these examples is to point out that elite behaviour does not just mean *linguistic* behaviour. Recall that the various defenders of elitism differ as to which types of behaviour they take to be central to belief. First, as has already been discussed, there is the distinction between conscious judgments and overt linguistic behaviour. Second, one can distinguish simple assertion from reflective endorsement – that is, an assertion which is made on the basis of a deliberative process and such that one could offer *reasons* in support of it. Perhaps if we look at a broader range of phenomena, a connection between elite behaviour and evidence will emerge.

It might be suggested that only one type of elite behaviour is sensitive to evidence in the requisite sense and that it alone is the mark of belief. Alternatively one might claim that the various types of behaviour tend to come as a package. Though any of these taken in isolation might not seem to have a special kind of evidence sensitivity, taken together they do. I think, though, that neither of these strategies can succeed. None of the types of behaviour have a sufficiently strong connection with evidence and they can all come apart from each other. Turning first to reflective endorsement, consider the following example:

*Malcolm the snap-judger:* Malcolm is a successful businessman who each day has to make dozens of assertions about people's character, fitness for a particular job, susceptibility to particular forms of persuasion etc. on the basis of quick decisions. If he were to reflect on any particular assertion he had made, he would often not endorse it; his on balance judgment would go against his unreflective assertion. However, at least in the sphere of business decisions, Malcolm's snap-assertions are more accurate than his reflective judgements.

This shows how our sincere assertions can come apart from what we reflectively endorse; and that, at least some of the time, what we say unreflectively can be more sensitive to the evidence than what we are disposed to endorse after deliberation. There is empirical evidence backing up the claim that cases like this are quite commonplace. For example, Halberstadt and Levine (1999) asked a group of basketball experts to predict the outcome of a basketball game: half were asked to give and analyse reasons before making their prediction and half were not; those who had to give reasons were less accurate on the whole than those who did not do so. Similarly, Wilson *et al* (1984) asked subjects who were in a relationship to predict how long it would last, with half giving reasons and half not. Again, it turned out that the predictions of the control group were significantly more accurate.

Of course there are also times when judgments arrived at reflectively are better attuned to the evidence than unreflective assertions – for example, when the subject matter of the assertion is some sophisticated area of science. The correct conclusion is that both types of behaviour are *imperfectly* sensitive to the evidence, and which fares better varies depending on the circumstances.

The next thing to consider is the relationship between assertion and conscious judgment. It might seem that these two are tricky to pull apart since we are almost always aware of what we are saying. Perhaps there are cases of talking on complete autopilot where we are totally oblivious of the words coming out of our mouth, but the separation here is shallow – we could quite easily become conscious of what we were saying by refocusing our attention. However, this only demonstrates a correlation between assertion and *linguistic* conscious judgment – internal ‘sayings to oneself’. There are other conscious occurrences that have a claim to being a manifestation of belief and that do come apart from what we say. Consider the following case:

*Maya the navigator:* Maya lives in Brussels – a city with winding disorderly streets. She is able to navigate it effectively by consulting a mental image she has of the city’s layout. This is a process that occurs reflectively and that she endorses as reliable. However, much of the information that she draws upon, she would be

unable to put into words: she could not describe the features that she is aware of when visualising the city.

This illustrates what should be an uncontroversial point: we sometimes have conscious representations – for example mental images – whose content we are unable to fully articulate. These mental images are, moreover, evidence sensitive: Maya’s mental map is formed, and can be updated, as a result of her perceptions of Brussels. Moreover, her verbal behaviour is not as sensitive to this perceptual evidence, since she cannot articulate such a rich description of the city. It should be noted again, though, that in *some* respects linguistic behaviour will do better at responding to the evidence. For example, if Maya receives a bunch of *linguistic* evidence about the city’s layout in an Urban Studies course, she will be able to make assertions about the ratio of public space to private space, the percentage of green space etc. And she will not be able to alter her mental image of the city in a way that encodes this information. So again, conscious states and verbal behaviour are both partially sensitive to the evidence, and which fares better depends on the circumstances.<sup>18</sup>

#### **4.3 Evidence and Non-Elite States**

We can conclude that elite states are sensitive to evidence in an imperfect and complex way. There’s no straightforward method for stating in what ways such behaviour is evidence sensitive; one has to make reference to the type of evidence in question and the circumstances in which the behaviour is elicited. This means that the only plausible evidence based constraint on belief is the minimal E4.

This level of evidence sensitivity is also exhibited by non-elite states, however. Though such states are often not sensitive to linguistic testimony, they often are sensitive to perceptual evidence. Consider, for example, Bob the baseball player: his disposition to move to the appropriate location to catch the ball is sensitive to perceptual evidence

---

<sup>18</sup> As variant on this case we can consider auditory rather than visual imagination. For example, a jazz musician without formal training might be able to imagine how various motifs would sound, and which would work in a given circumstance – this might guide her playing though she would be unable to verbally articulate what she was imagining. Moreover, the imaginative abilities might be updated as she heard new performances.

regarding the ball's trajectory through the air. Similar things could be said about much action in sport. More generally, we have dispositions to navigate environments correctly that we are not conscious of. For example, there are many buildings that I have visited just a few times and that I am able to find my way around when I return, but that I am unable to visualise. Moreover, in a familiar environment I will be able to reach for door handles, light switches and elevator buttons without looking, despite being unable to picture their location. These dispositions too are shaped by perceptual evidence.

A crucial point is that even the non-elite states of the subjects in the original cases may display this level of evidence sensitivity. Note that the examples as presented do not specify whether the non-elite states are evidence sensitive. However, there are plausible ways of filling in the details that make this the case. First take the trembling and sweating of the subject on the skywalk. This behaviour is clearly sensitive to perceptual evidence since it's seeing through the glass bridge that causes it. Moreover, it might be that if the subject were to vigorously jump up and down, throw herself against the walls etc. and see that everything still held firm, her trembling would cease – sometimes this kind of visceral demonstration is an effective technique for getting over unnecessary fear. Second, Ben's disposition to drive along the wrong road is sensitive to visual evidence – he'll turn back when he sees signs for diversions around the bridge; moreover, he'll lose the disposition (we may stipulate) once he's made the mistake a few times. Delayed sensitivity to the evidence is enough to satisfy E4: we've seen that elite states may not be *immediately* sensitive to evidence either, as with Thomas the slow learner.

The case of implicit bias is more complicated since the non-elite behaviour in question is manifested in a vast range of situations and, moreover, there is much uncertainty over how it is formed and how it can be changed. There does seem to be evidence that a subject's behaviour will become less biased if she is exposed to people who do not fit with the stereotype. For example, Dasgupta and Greenwald (2001) found that exposing subjects to admired black exemplars and disliked white exemplars, significantly reduced implicit bias behaviour. And Shook and Fazio (2008) found that white students who shared a dorm room with a black student, saw a reduction in implicit bias over the time they spent together. This suggests that even though testimony that black people are not inferior does not influence implicit bias behaviour, more direct

forms of evidence may do – such as witnessing up close a black person who does not conform to the stereotype. This is comparable to the partial evidence sensitivity of Rich the meat eater’s linguistic dispositions.<sup>19</sup>

Together, these considerations show that the fact that beliefs are evidence sensitive does not speak in favour of elitism. In fact, an examination of the relationship between various belief-like states and evidence points in the opposite direction since it reveals important similarities between elite and non-elite states.

## 5. Inferential Integration

The second argument for elitism appeals to the connection between belief and *inferential integration*. It seems a plausible general principle that beliefs are inferentially integrated – in a sense to be spelled out below – and this might be thought to support elitism. The canonical presentation of this idea is given by Stich (1978), who says: ‘a person’s body of beliefs forms an elaborate and interconnected network with a vast number of potential inference patterns leading from every belief to almost any other.’<sup>20</sup>

He claims, moreover, that the states which have this property are exactly those states that we have ‘conscious access’ to – where the mark of conscious access to a state with the content *p* is the ability to *say* that *p* when asked, or to consciously judge that *p* when considering the question.<sup>21</sup> The elitist, therefore, could make the following argument:

1. All beliefs are inferentially integrated.
2. Only elite states are inferentially integrated.
- C. Only elite states are beliefs.<sup>22</sup>

There certainly seems something right about the idea that beliefs must be inferentially integrated (premise 1): intuitively, beliefs are the kinds of thing that can result from and

---

<sup>19</sup> Schwitzgebel (2010) pp. 539-541 makes a similar point

<sup>20</sup> Stich (1978) pp. 42-43.

<sup>21</sup> *ibid* pp. 40-41.

<sup>22</sup> I think that Stich himself is working with a slightly different dialectic. He takes it to be intuitively obvious that all beliefs are elite states (mistakenly, in my view) and wants to argue that inferential integration is a property distinctive of them – and thus that our concept of belief is a category of theoretical interest.

produce inferences. Moreover, if behaviour results from an inferentially integrated network of states, so that a large body of information is brought to bear on it, this seems like a good reason to think of it as intelligent, rather than a reflexive response to a given stimuli. And as was mentioned above: beliefs produce intelligent behaviour.<sup>23</sup>

Stich motivates the claim that only elite states are inferentially integrated (premise 2) by example. Suppose that we cognitively encode certain propositions about the syntax of our language – let *p* be such a proposition (e.g. one providing some constraint on anaphora binding). This would be a paradigm case of a non-elite state. Stich notes that a subject might, as a result of her research in linguistics, believe that if *p* holds then Chomsky is mistaken – she would assert such a thing in conversation – and she might also cognitively encode *p*, employing it in her processing of language. However, she would not be in a position to conclude that Chomsky is mistaken: if you asked her whether Chomsky was mistaken, she'd say she didn't know, perhaps adding that she didn't know whether *p* held.<sup>24</sup> This shows that the subject's representation of *p* is not inferentially integrated with her belief that if *p* then Chomsky is mistaken. On the other hand, if a subject was disposed to *assert* both '*p*', and 'if *p* then Chomsky is mistaken' then she would presumably be able to conclude that Chomsky is mistaken. I think that Stich is right about this example but I don't think all non-elite states are like the cognitive encoding of linguistic rules.

Before assessing the argument, though, a clarification is in order about how to think about inference. Stich's main example of it is *modus ponens*. This might suggest that inference is a relation that can hold only between entities with a sentential structure, since the most natural way to think of *modus ponens* is as a relation between sentences (you need a conditional, after all). However, it's clear that our ordinary conception of inference doesn't have this requirement – when it comes to imagistic reasoning for example. Suppose I have a Klee painting that I want to hang in my kitchen; I may visualise both the room's layout and the painting in order to form a belief as to whether it is too big to fit between the fridge and the oven. This process seems a perfectly good

---

<sup>23</sup> A related argument is Fodor's (1983) influential claim that beliefs are states within the *central system* – as opposed to informationally encapsulated modules. I stick with Stich's formulation since it requires less technical machinery.

<sup>24</sup> See *ibid* p 44

example of an inference – I draw upon information I have about the painting and my kitchen to see whether a particular action is possible. Moreover, it's an *open question* whether such processes involve analogue as well as sentential representations – the imagery debate is not yet resolved.<sup>25</sup> We are happy to classify processes as inferences without presuming they involve only sentential entities.

I think, therefore, that we should work with a minimal conception of inference which makes no assumptions about the types of representation it involves – sentential, analogue, or whatever. I do not propose to offer an analysis of the concept, since I think we have a decent intuitive grip on. In what follows, I will work with this intuitive understanding.

Now returning to the argument, it's important to distinguish two theses about inferential integration – strong and weak.

*Strong integration:* A state is a belief only if it is inferentially integrated with *all* other beliefs.

*Weak integration:* A state is a belief only if it is inferentially integrated with a *sufficient number* of other beliefs (but not necessarily all of them).

My strategy will be to argue that the strong integration thesis is false, and that the weak integration thesis, though plausible, does not favour elitism since non-elite states also satisfy the conditions it sets.<sup>26</sup>

Strong integration is too strong since not even paradigm cases of belief by the elitist's lights meet it – some elite states cannot be integrated with others. First, it's clear that the information manifested in linguistic behaviour cannot always be integrated with that represented in conscious mental images. Recall the example of Maya the navigator: she cannot verbally articulate the information represented in her mental map, and she cannot use the theoretical information she asserts to modify the map. In fact, experiments

---

<sup>25</sup> See, e.g., Tye (2000).

<sup>26</sup> So in the argument above, if we work with strong integration then premise 1 is false, while if we work with weak integration premise 2 is false – either way the argument is unsound.

have shown that in certain cases, attempts to articulate the contents of visual memory – e.g., to describe a remembered face – actually degrade it.<sup>27</sup>

The elitist might respond by denying that mental images are manifestations of belief – that all belief must be manifested in *linguistic* behaviour. As well as leading to an unattractive picture of belief, this move fails to rescue strong integration. For example, sometimes whether we are able to provide a given piece of information verbally depends on what question we are asked. Suppose you are asked ‘is there a four letter English word ending E-N-Y?’ – you might well not be able to answer. However if you are asked ‘how do you spell *deny*?’ you will correctly answer ‘D-E-N-Y’. So you can access the information ‘the English word *deny* is spelt D-E-N-Y’. Moreover, you might well be able to give examples of words ending, A-N-Y, I-N-Y, and O-N-Y on demand (‘many’, ‘tiny’, ‘pony’). And presumably you could say that ‘puny’ is spelt P-U-N-Y. So you have information that together entails that for every vowel x, there is an English word that ends x-N-Y. However, (prior to reading this paragraph) you were not capable of putting this information together to draw this inference.<sup>28</sup>

In response to this, the defender of strong integration might appeal to *cognitive architecture*. My examples show that sometimes subjects are not, intuitively speaking, able to *bring together in inference* certain of their elite states. One could argue that this is not how the claim that ‘all beliefs can be inferentially integrated with all other beliefs’ is to be interpreted. Instead what might be relevant is some sort of *in principle* accessibility – that there are no barriers in virtue of cognitive architecture, only ‘performance limitations’.

This response fails, however, since the sophistication constraint comes from a pragmatic account of belief. For practical purposes, what matters is whether beliefs can be integrated in everyday context, not what holds as a matter of cognitive architecture. Thus, I will turn to *weak integration*. My contention is that many non-elite states are weakly integrated. One of the most compelling examples of this is, I think, behaviour in sport. Consider the following case:

---

<sup>27</sup> See Schooler and Engstler-Schooler (1990).

<sup>28</sup> The original trick regarding ‘deny’ is found in Powers (1978).

*Federer:* In the 2006 Wimbledon final between Federer and Nadal, the following rally takes place. Federer hits three backhands down the line, causing Nadal to stay planted on that side – lulled into a false sense of security. He then hits two hard shots to the opposite side, taking advantage of Nadal’s flat-footedness and forcing him to scramble. This leads Nadal to play a weak short shot, allowing Federer to return at a very sharp angle so that he wins the point.<sup>29</sup>

In this example, Federer acts in an intelligent way, executing a complex and difficult plan to win the point. It is not habitual or reflexive behaviour, since it is tailored to Nadal’s specific abilities – against other players Federer could have tried to win the point sooner, but Nadal is exceptionally quick. It’s also open to modification depending on exactly what Nadal throws back, and when and how he gets wrong-footed. This strongly suggests it is the product of a network of inferentially integrated states since it is sensitive to information received from a variety of sources over an extended period of time – background knowledge of Nadal and of Federer’s own abilities, and perceptual information about what the ball is doing, the court conditions etc. It is not, though, the product of a conscious process of deliberation; it happens far too fast for that.

There is also this level of inferential integration in the cases mentioned in previous sections. Bob the baseball player’s disposition to move to catch the ball may have been calibrated on the basis of a wide variety of information such as his visual perception of the ball, his sense of the wind strength and direction, and his kinaesthetic sense of the state of his own body (how fast he can run, dive etc.). Ben is not only disposed to drive on the old route to work, but will take his dry cleaning even though he only passes the dry cleaners on the old route; and if he fancies a bagel, he will change the way he is driving to go past the deli etc. Recall also the example of a building I have visited a few times and can remember my way around as I go – though I can’t articulate or visualize its layout. When I enter it, I will alter my behaviour depending on what I aim to achieve there, and also if I get new perceptual evidence that the layout has changed. With implicit bias the characteristic behaviour manifests itself in a variety of ways – body language, speech, workplace decisions etc. It also appears to draw on background information in its

---

<sup>29</sup> This is a paraphrase of the commentary given by David Foster Wallace (2006).

activation – hiring decisions are unconsciously influenced by the racial connotations of the name on the CV and so would appear to be mediated by background beliefs about which names are typical of which races.<sup>30</sup>

I conclude that the appeal to inferential integration to establish elitism fails.<sup>31</sup> In the next section I'll briefly look at where this leaves us when it comes to understanding belief.

## 6. Conclusion

This shows that both elite and non-elite states can be instances of belief – they both fit with the practical role of the action account. Note that this confirms that the action account supports intellectualism for know-how and is in tension with accessibility constraints on knowledge and belief.

Rejecting elitism raises further questions about cases of conflicting behaviour, however. It suggests normal subjects have a vast range of conflicting *beliefs* – it's a tricky business to determine how we should ascribe beliefs in such cases in a manner that fits with the practical role of belief. This is the topic of the next chapter.

---

<sup>30</sup> See Mandelbaum (2012) for further argument that implicit bias must involve inferential reasoning.

<sup>31</sup> As I mentioned above, Davies (1989) – among others – has argued that one better captures the sophistication of belief by appeal to the generality constraint, rather than inferential integration. I think if one were to try to run the argument in these terms, my objection would still apply: elite states only satisfy the generality constraint partially and non-elite states do so too. Spelling out this argument, however, would require setting everything up in terms of the contentious theory of concepts the generality constraint presupposes, so I won't pursue the matter here.

[blank by design]

## Chapter 4

# Task-Indexed Belief

### 1. Introduction

We have endorsed a permissive account of belief, rejecting elitism in cases of conflicting behaviour. The question remains as to how exactly we ascribe beliefs in such cases. In this chapter, I will argue that, due to the demands of pragmatism, belief must be indexed to *tasks*. That is, whereas belief is traditionally thought of as a two-place relation between a subject and a proposition, on my view it is a three-place relation between subject, proposition, and task. No simple yes/no answer to the question ‘does S believe that p?’ is acceptable in cases of conflicting behaviour – instead one needs to say ‘S believes that p relative to one task but doesn’t relative to another’.<sup>1</sup>

My central motivation for this claim is that belief’s action-guiding role is too *complex* to be modelled by a two-place account. Discussion of conflicting behaviour has typically focused on the conflict between verbal behaviour and conscious judgment on the one hand, and non-verbal unconscious behaviour on the other. However, the examples in the previous chapter, along with some new ones, show that the situation is not so simple – our behaviour can come into conflict along multiple dimensions. I’ll argue that two-place accounts are not capable of modelling such cases adequately, while my indexed account is designed to do so.

The pragmatist methodology entails that, all things being equal, the account of belief that can model these cases in an informative manner is the correct one – thus providing an argument for indexed belief. This is a significant result since, as I’ll show in the final

---

<sup>1</sup> The idea of postulating a third parameter for belief is not new; however, most previous discussions of this have been concerned with traditional Frege cases, which I think is an independent issue. I’ll discuss this further below.

sections, it leads to some surprising conclusions about belief attribution, modes of presentation, and the knowledge argument.

## 2. Traditional Cases of Conflict

Traditional cases of conflicting behaviour are those in which a subject's elite behaviour suggests belief in one proposition while her non-elite behaviour suggests belief in its negation. We discussed a number of these in the previous chapter:

*Roadworks*: In which Ben has elite behaviour suggesting he believes that the bridge is closed, while his non-elite behaviour suggests a belief that it's open.

*Implicit Bias*: In which the implicitly biased subject's elite behaviour indicates belief that black people are equally trustworthy and deserve equal treatment to white people, while the unconscious behaviour indicates a belief that black people are inferior.

*Federer*: In which Roger has non-elite behaviour that indicates a belief that his opponent is wrong-footed while his elite behaviour indicates he does not believe this.

How should we ascribe beliefs in these cases? If one takes 'does S believe that p?' to be a straightforward yes/no question, then there are three natural responses:

1. *Elitism* (just as a reminder): In all cases the subject believes what is indicated by her elite behaviour. Ben believes that the bridge is closed; the implicitly biased subject believes that all races are equal. Roger doesn't believe her opponent is wrong-footed.
2. *Anti-Elitism*: In all cases the subject believes what is suggested by her non-elite behaviour. Ben believes that the bridge is open etc.

3. *Contradictory Belief*: In all cases the subject both believes that p and believes that not-p for the relevant proposition. (Both elite and non-elite states are beliefs.)<sup>2</sup>

We have seen, though, that the states underlying both kinds of behaviour are sufficiently sophisticated to qualify as *beliefs* – so elitism and anti-elitism can be ruled out. The options are not exhausted though, since one might think that cases of conflicting behaviour violate the normal conditions for belief ascription, meaning there's no determinate answer as to what the subject believes, leading to another possibility:

4. *In-Between Belief*: In all cases it's indeterminate what the subject believes. Or as an alternative formulation: the subject is *in between* believing that p and believing that not-p.<sup>3</sup>

Finally, one might take seriously the talk of elite and non-elite states – ruling them two different *kinds* of state. And thus claim that 'belief' is strictly speaking ambiguous, giving us the following view:

5. *Two Statism*: The term 'belief' is ambiguous – there are two kinds of state, belief1 and belief2, such that elite behaviour is caused by one and non-elite behaviour by the other.

All of these views have been taken seriously, and sustained arguments have been offered in their favour.<sup>4</sup> At this point, I have only distinguished views in terms of *first-order*

---

<sup>2</sup> Strictly speaking, there is a fourth possibility: in all the cases the subject neither believes that p nor believes that not p. This would seem to make these cases of suspension of judgment. I will not consider this view since it does not seem plausible and has not, as far as I know, been defended in the literature.

<sup>3</sup> One might think the Bayesian framework could be of help in making precise the notion of in-between belief – presumably equating it with middling credence. However, this is a non-starter. The subjects have elite behaviour suggesting very high confidence in a proposition (Ben would take bets at high odds that the bridge was closed) and non-elite behaviour suggesting high confidence in its negation – this is nothing like a typical case of 0.5 credence.

<sup>4</sup> See chapter 3 for details on supporters of elitism. Gertler (2011) defends anti-elitism. Shoemaker (2009) defends contradictory belief. Schwitzgebel (2010) defends in-between belief. Frankish (2004) defends two-statism. A difficult case to classify is Gendler (2008a+b). She is officially an elitist since she is only willing

doxastic attitudes towards the relevant proposition. Such accounts can be elaborated in various ways – providing details about the *higher*-order beliefs and other kinds of mental state of the subjects. In section 3.3, I discuss the extent to which these details help respond to objections.

Each position gives a fairly neat account of the traditional cases: however, such examples are only the tip of the iceberg. Conflict between elite and non-elite states is one instance of a whole *family* of cases of conflicting behaviour. An adequate account of belief should be able to deal with all of these appropriately. I will show that none of the two-place accounts can do so, motivating a move to task-indexed belief.

### 3. The Problem of Inversions

In this section, I present the range of examples that prove problematic for the various two-place accounts. The problem arises because we have subjects who have very different mental states, but cannot be distinguished by the two-place accounts. Since the examples involve taking one case and inverting features of it, I call this objection ‘the problem of inversions’. Not being able to track the differences between such agents is a disadvantage for an account of belief; I’ll explain in more detail in section 5 how pragmatism makes this a serious problem.

#### 3.1 In-between Belief and Contradictory Belief

First we’ll look at how the problem arises for in-between belief and contradictory belief. Take the example of Ben in *Roadworks*: his linguistic behaviour suggests that he believes that the bridge is closed while his non-linguistic behaviour suggests that he believes it is open. Now according to the in-between belief account, Ben is in a state of in-between belief with regard to the proposition that the bridge is open; according to the contradictory belief account he both believes that it is open and believes that it is closed. According to both accounts, this is all that’s to be said at the doxastic level.

---

to refer to elite states as ‘beliefs’, however she does postulate a second belief-like state (‘alief’) which she thinks is of theoretical significance – therefore, showing some affinity with two-statism.

However, consider Beth whose behaviour is an ‘inversion’ of Ben’s. She is disposed to assert that the bridge is open and use the claim as a premise in verbal reasoning; but she will take the detour avoiding the bridge when driving to work etc. Again, the in-between belief account says Beth has an in-between belief and the contradictory belief account says she has contradictory beliefs. So according to both of these accounts, Ben and Beth are doxastically indistinguishable – they can’t be distinguished in terms of what they believe. However, there are important differences between these two subjects: there will be situations in which the difference between Beth and Ben’s pattern of behaviour is crucial. In the one case we might be concerned with whether a given subject will be late to the meeting, in the other with what directions they gave to our friend. Moreover, it’s plausible that these are the kinds of things, practically speaking, that we want belief ascriptions to distinguish. Therefore, these accounts do not allow us to communicate everything we want belief ascriptions to communicate.<sup>5</sup>

Similarly, the implicitly biased subject and an inverted subject who is explicitly racist but unconsciously egalitarian will hold the same beliefs, according to both of these accounts. Indeed an approximate case of this inversion is Huck Finn, who avows that slavery is just, but refuses to give an escaped slave back to his owners.<sup>6</sup> As the sizable (and distinct) literatures on Huck Finn and implicit bias demonstrate, these are cases we have a great interest in understanding and distinguishing.

Also, consider a sports journalist who can articulate information about sophisticated tennis strategy (we can stipulate that this is the information Roger draws upon while playing), but who is unable to play tennis skilfully. Again the difference between these two subjects couldn’t be starker, yet the in-between belief and contradictory belief accounts are unable to distinguish them.

These inversions do not present a direct problem for elitism and anti-elitism (previous objections notwithstanding). According to elitism Ben simply believes that the bridge is closed and Beth simply believes that it is open (and *vice-versa* for anti-elitism), so their doxastic states differ. However these views face a closely related problem: they cannot distinguish conflicted subjects from their non-conflicted counterparts. For example,

---

<sup>5</sup> Elga and Rayo (ms) also note this problem for these accounts.

<sup>6</sup> See Arpaly (2003).

elitism classifies both the implicitly biased subject and the subject who is both implicitly and explicitly egalitarian as simply believing that all races are equal.

### 3.2 Possible Responses

I want to pause and consider how defenders of such views might respond to this objection. The general strategy would be to claim that though these subjects can't be distinguished in terms of their first-order *beliefs*, they can be distinguished with some additional information.

A first-pass attempt at this would be to simply describe the different patterns of behaviour. For example, the contradictory belief theorist could say that it's Ben's belief that the bridge is closed that guides his elite behaviour and his belief that it's open that guides his non-elite behaviour (and *vice-versa* for Beth). Similarly, the in-between belief theorist could specify how the 'bridge-open side' of Ben and Beth's in-between beliefs shows up in different kinds of dispositions. Further, the elitist could say that the implicitly biased subject possesses certain implicit dispositions that the unconflicted subject lacks.

However, to do this is to give up on offering a *systematic* account of the cases – resorting instead to brute specification of behavioural dispositions. I think that an account that can avoid doing this is to be preferred, since this method misses out on the practical benefits of belief ascription.

A more sophisticated response is to distinguish the subjects by means of some different kind of mental state. One option would be to appeal to higher-order beliefs: that is, though subjects and their inverts do not differ in first-order beliefs, they differ in what they *believe they believe*. For example, the implicitly biased subject believes that she believes that all races are equal, while Huck Finn believes that he believes that black people are inferior.

In order to assess the force of this response, it's important to remember that my objection only requires that there are *some* examples that can't be distinguished by the traditional account, not that they can't distinguish any of them. There may be some ways of filling in details of these examples so that higher-order beliefs play the required role,

but there are also, I think, cases where this is not so. For example, we can first imagine that the implicitly biased subject reflects on the fact that she has examined the evidence on racial characteristics and that she has many times asserted that all races are equal, so she judges that she believes that all races are equal. It's plausible that such a subject believes that she believes that all races are equal.

However, we could also imagine an implicitly biased subject who asserts that all races are equal but who has never explicitly considered the question of what she *believes* about racial equality. It's much less plausible to think that such a subject possesses the requisite higher order belief – and thus the problem of inversions applies in this case.<sup>7</sup>

There are other ways one might make this kind of response. For example, the elitist might claim that Ben differs from a non-conflicted subject, not in having another propositional attitude, but in being prone to temporarily *forget* that the bridge is closed, while the belief of the regular subject is *stable*. Similarly, the anti-elitist might claim that the implicitly biased subject differs from the explicitly racist subject only in that she's in *denial* about what she believes.<sup>8</sup>

I think such responses will be unsuccessful, though, for the same reasons as with the higher-order belief response. Though they may apply to certain specifications of certain examples, they will not be able to provide a comprehensive solution. I will not attempt to go through each response exhaustively, since I think the discussion so far makes it plausible that no attempt can succeed (Gendler (2008a) provides a systematic and persuasive discussion of why accounts along these lines cannot succeed). Instead I will look at whether an alternative account fares any better.

### 3.4 A Two State Solution?

---

<sup>7</sup> One might respond by claiming that elite dispositions always come with higher order beliefs, while non-elite dispositions never do – thus one can always appeal to such states to distinguish subjects and invert. Such a claim is highly controversial, however. Also, if one accepts it, one will not be able to use higher-order beliefs to explain cases in which different kinds of elite behavior conflict with each other – as discussed in §3.3 below.

<sup>8</sup> Another possibility is that the elitist could say the implicitly biased subject possesses implicit attitudes or, following Gendler (2008a+b), an 'alief', that the unconflicted subject lacks. However, such a response is not satisfactory. If the additional state is representational, as I think aliefs are meant to be, the view will essentially be a variant on two-statism, and so will face the objections discussed below. If it isn't, it will amount to a brute specification of dispositions and so be unsatisfactory for the reasons mentioned above.

At this point, two-statism might look rather attractive. Such a view, recall, proposes that where we talk loosely of ‘belief’ there are really two types of state – the one underlying elite behaviour (call it *belief1*) and the other underlying non-elite behaviour (call it *belief2*). This gives us the machinery to distinguish the cases above. Ben believes<sub>1</sub> that the bridge is closed and believes<sub>2</sub> that it’s open, while Beth believes<sub>1</sub> that it’s open and believes<sub>2</sub> that it’s closed. Similarly, the implicitly biased subject believes<sub>1</sub> that all races are equal and believes<sub>2</sub> that black people are inferior and *vice versa* for Huck Finn.

However, though it works for these particular cases, two-statism does not solve the problem in general. There are two ways the problem arises. The first stems from the fact that there are multiple kinds of elite behaviour, typically taken to come as a package, but that in fact often come into conflict. The second is that whether a state is manifested may depend upon the circumstances or the purposes of the subject’s action. Importantly, I think the sheer *variety* of problem cases shows that no tweaking of the two-state proposal will be sufficient – nor indeed opting for a three-state solution! It is this that motivates moving to an indexed view since doing so buys us a great deal more expressive power.

To get the problem going, recall from chapter 3 that elite behaviour includes both *linguistic* behaviour, and *conscious judgments* – the latter are internal phenomenal events, including ‘sayings to oneself’, as well as non-linguistic representations such as mental images.<sup>9</sup> Further, there is a distinction between simple *sincere assertion*, and *reflective endorsement* – that is, an assertion which is made on the basis of a deliberative process and such that one could offer *reasons* in support of it. Elite behaviour is taken to be a unified category by many theorists, but as we’ve discussed, this assumption runs into problems. Elite behaviour *can* come into conflict, which both undermines the significance of the belief<sub>1</sub>/belief<sub>2</sub> distinction, and allows the problem of inversions to re-emerge. We can illustrate this using two of the other examples from the previous chapter:

*Maya the navigator*: Maya is able to navigate Brussels effectively by consulting a mental image she has of the city’s layout. However, she could not describe the features that she is aware of when visualising the city.

---

<sup>9</sup> That is, states that are not *explicitly* linguistic in form – I do not wish to take a stand on the ‘imagery debate’. See Pylyshyn (1981).

We can elaborate on this example so that it leads to an inversion problem for the two-statist. For example, Maya might take an Urban Studies course which gives her all sorts of information about Brussels. She will then be able to make assertions about the ratio of public space to private space, the percentage of green space, etc. But she will not be able to alter her mental image of the city in a way that encodes this information.

To keep the case simple, suppose that Maya's mental image has as part of its content that the court is to the east of the railway line, while she is disposed to assert that it lies to the west, having been taught that all government buildings are on this side. Since these are both forms of elite behaviour, the two-statist must say that Maya both believes<sup>1</sup> that the court is to the east of the railway line and believes<sup>1</sup> that it is to the west.<sup>10</sup> However, one could imagine an invert of Maya who asserts that the court is to the east and has a mental image with the content that it is to the west. Two-statism will have to ascribe the same mental states to this subject as to Maya.

Now consider another familiar example:

*Malcolm the snap-judger:* Malcolm is a successful businessman who each day has to make dozens of assertions on the basis of quick decisions. If he were to reflect on any particular assertion he had made, he would often not endorse it.

This case again leads to inversions. Let's imagine Malcolm is disposed to give confident snap assertions on a range of business matters, while he is disposed to remain ambivalent when considering the propositions in a reflective manner. Then the two-statist would have to model Malcolm as both believing<sup>1</sup> various propositions about business and also suspending judgment about these claims. We can easily imagine a person, though, who is unwilling to commit initially to any proposition but will tend to confidently endorse claims after reflection (many philosophers fall into this camp). Such a person would be an inversion of Malcolm and the two-statist would be unable to distinguish their mental states.

---

<sup>10</sup> Alternatively, one might say she in-between believes<sup>1</sup> both that it is to the east and the west. The problem will remain.

One might argue that Malcolm is not a subject with conflicting attitudes, but someone who is disposed to *change his mind*. However, though this might be right for judgments about particular people – for example he first forms the belief that Janet is a good negotiator then abandons it on reflection – it cannot capture what’s going on with the *general rules* he follows when forming these particular beliefs. Suppose Malcolm has a stable disposition to make snap judgments that people with confident body language are good negotiators while he stably rejects this generalization on reflection. These both seem like they involve stable beliefs, by the two-statist’s lights, and we can imagine an inverted subject here too.

A second strain of problem for the two-statist arises from the fact that conflict occurs not just based on different kinds of behaviour but on the different *circumstances* in which it’s elicited, or purposes for which it’s required. Take the following case:

*Jane the sporadically biased:* Jane is a university professor who avows the equality of intelligence between races when speaking in public, whilst her thoughts in private moments are prone to register reservations. Moreover, she thinks favourably of black people she meets at academic conferences (and treats them fairly) whilst the thoughts she has when passing black people on the street at night are unduly suspicious.

Here, the conflict is between behaviour in a comfortable setting and in a threatening setting. One could imagine an inversion of Jane who was biased in academic settings but not in other contexts – which again, the two-state account is unable to accommodate adequately. There are familiar cases, analogous to this, involving how people’s judgements vary depending on the company they’re in: their cool new friends, or their old school friends, or their family etc.<sup>11</sup>

Cases of conflict that involve only non-elite behaviour also arise:

*Defensive Josie:* Josie is a defensively talented but offensively weak basketball player. While defending, she is able to spot dangerous situations: she knows who to

---

<sup>11</sup> Sometimes people might be faking it to fit in, but some of the time it happens with complete sincerity.

box out to avert the danger, what the offense will try to do to stop her and how she can counteract their moves. However, when her team is attacking, she is incapable of making the correct decisions.

In this case, Josie is only able to draw upon information concerning what should be done in basketball plays in defensive situations. For example, consider the claim that if an off-ball screen is set, the nearest defender should offer help: when defending Josie is able to offer such help, but when attacking she does not anticipate it coming. And, of course, we could imagine her invert (*Offensive Josie*), an offensively gifted and defensively poor player – again, the two-state account would be unable to distinguish the two.

Taken together, these examples strongly suggest that no two-place account of belief can handle all the cases of conflicting behaviour. There are multiple dimensions of variation in how belief-like states guide behaviour, and thus multiple dimensions of conflict that need to be adequately described. Two-state accounts just don't have the power to do this, no matter how you finesse the details.

#### **4. Introducing Indexed Belief**

The discussion and examples above show that we are subjects with states that guide behaviour in some ways but not others. Moreover, which 'some' and which 'others' can be of great importance – so we want our account of belief to be able to capture this, in as systematic a way as possible. My strategy for doing this is *indexing* belief to the parameters of variation. The framework here is similar to that involved in so-called *fragmentation* accounts of belief – introduced by Stalnaker (1984) and Lewis (1982) and more recently pursued by Elga and Rayo (ms) and Egan (2008).<sup>12</sup>

There are many different ways belief can manifest itself: in linguistic behaviour; in mental imagery; in unconscious behaviour; in a comfortable setting or in a dangerous setting; playing defence or offense in basketball. There are multiple dimensions along

---

<sup>12</sup> Though the tools are similar the application is quite different. Stalnaker and Lewis introduced the notion primarily to model cases of mathematical ignorance and, more generally, differing attitudes towards necessarily equivalent propositions. Elga and Rayo aim to explain a whole range of cases including limited recall capacities and deductive abilities. Though I see these projects as in many ways complementary I am not committed to their conclusions.

which the conditions for belief manifestation can vary: in some cases it concerns the kind of behaviour (e.g., verbal); in others, the circumstances of the subject (e.g., at a conference); in others still, the purposes of the action the subject is performing (e.g., playing defence).

Speaking generally we have various categories of action – a token act will belong to a category if it possesses a particular property, whether that concerns the circumstance, purpose, or nature of the act. I propose to use the term ‘task’ to refer to these categories of action. It should be noted that the expression is being used stipulatively: natural examples, such as participating in a conversation, will count as tasks in my sense; but so will less intuitive categories like performing actions in a threatening situation.

We need to individuate belief states in terms of the tasks they guide. Indexing is the most direct way of doing this. Roughly, a subject *S* believes that *p* relative to a task *t* if *S* has a belief-like state with content *p*, which guides the aspects of her behaviour covered by *t*. This means that instead of the single question ‘does *S* believe that *p*?’ we have multiple questions of the form ‘does *S* believe that *p* relative to *t*?’

If we can help ourselves to talk of indexed belief, we will be able to distinguish the subjects and inverters from the previous section. For example, (roughly) Ben believes that the bridge is open relative to practical tasks but believes that the bridge is closed relative to verbal tasks. Conversely, Beth believes that the bridge is closed relative to practical tasks but believes that the bridge is open relative to verbal tasks. This means that we can distinguish Beth and Ben’s doxastic states with indexed belief. Similarly, the implicitly biased subject believes that all races are equal relative to conscious tasks but not relative to (certain sorts of) practical tasks; conversely, Huck Finn believes that black people deserve to be subjugated relative to verbal tasks but not relative to practical tasks. Finally, Maya believes various things about the layout of Brussels relative to tasks involving visual imagination and navigation but not relative to verbal tasks – and *vice versa* for her invert. Similar things can be said for all the examples that have been discussed – for the sake of brevity, I won’t list them all.

This shows that indexed belief would be a great thing if we had it, since it would be able to overcome a problem faced by all the previously discussed accounts. To show that

this is a respectable strategy, I need to give an account of what holding a belief relative to a task amounts to, rather than relying on leading terminology.

This can be done by looking at the functional role of belief. The practical role of belief ascription aims to pick out states that guide behaviour in a sophisticated manner. In particular our beliefs guide our attempts to realize our desires and carry out our plans. On a minimal conception of this functional role, to have a belief that *p*, one must be disposed to behave in ways that would make sense if *p* held, given the desires one is attempting to satisfy. Or more precisely, one tends to act in ways that would realize one's desires, were it the case that *p*.<sup>13</sup> The examples show that this behaviour guidance doesn't apply in a straightforward way – for some tasks a subject will behave in ways that would make sense if *p* held, but for others she will not. One can add this wrinkle into the functional account as follows:

*Indexed Belief*: S believes that **p relative to *t*** only if S is in a state *X* such that:

*X* causes behaviour **included in the range of *t*** that would bring about her desires, were it the case that *p*. (Where *t* picks out a range of actions.)

To put it more informally, one believes that *p* relative to *t* only if one is in a state that causes one to behave in ways appropriate for the belief that *p*, when it comes to the behaviour covered by *t*. It's important to point out that I am not attempting to give a functionalist *definition* of belief here, but rather specifying some features the state must possess. Even if one gives up on the idea of a functionalist analysis, it's still highly plausible to think that functional constraints play a role in individuating beliefs. And *Indexed Belief* is part of this story.

The constraint fits with the examples we've been discussing. Ben has verbal behaviour that tends to realize his desires if the bridge is closed, since he asserts that it is and desires to give accurate testimony to others. His non-verbal behaviour satisfies his desire to get to work as quickly as possible if the bridge is open. So as wanted, Ben comes out as

---

<sup>13</sup> Few philosophers are willing to go so minimal as to take this condition to be both necessary and sufficient for believing that *p*. A possible exception is Marcus (1990).

believing that the bridge is closed relative to verbal tasks and as believing that it's open relative to practical tasks. The implicitly biased subject desires to treat people as they deserve, and to get the best candidates working in various positions. Her linguistic behaviour achieves this if black people are equal (e.g., her signing off on an affirmative action proposal). Her unconscious behaviour achieves this if they are inferior (micro aggressions are demeaning; disregarding *resumes* at a glance efficiently weeds out inferior candidates). Therefore, the implicitly biased subject believes that black people are equal relative to verbal tasks but that they are inferior relative to practical tasks.<sup>14</sup>

*Indexed Belief*, as stated, works with a minimal conception of belief's role in guiding behaviour – only making reference to its interaction with desires. It's thought by many philosophers that the real story must be more complicated. Accordingly, there are all sorts of ways to make this account richer – for example, making reference to other attitude states such as intentions and fears.<sup>15</sup> Or allowing that beliefs and desires come in degrees. One might also want to allow for structured or conceptual content – perhaps postulating inferential roles for various beliefs.<sup>16</sup> Further, one might want to build in some extra causal structure – for example, counterfactual independence between belief states and desire states.<sup>17</sup> At this point, I don't want to take a stand on which of these measures are appropriate – aside from my previous arguments that beliefs must be sufficiently inferentially integrated, and evidence sensitive. What matters is that whatever additions are made, the indexing aspect of functional role can be retained – thus the move I'm defending is compatible with a range of views in philosophy of mind.<sup>18</sup>

We've seen that *Indexed Belief* provides a solution to the problem of inversions. I've argued that this gives the view an advantage over its two-place rivals. The question remains as to how significant this advantage is. To assess this, we need to look more carefully at how this issue relates to the practical role of belief.

---

<sup>14</sup> Similar considerations hold for the other examples we've been discussing.

<sup>15</sup> This is to deny the Humean view that such attitude states can be *reduced* to beliefs and desires. See, e.g., Bratman (1987).

<sup>16</sup> See Peacocke (1992).

<sup>17</sup> See Braddon-Mitchell & Jackson (2007) ch 7.

<sup>18</sup> Another complication is that it is likely that *other* attitude states, such as desires, will be manifested in some circumstances but not others. Thus a line of argument similar to the one pursued here would suggest that desires must also be indexed to tasks.

## 4.1 Indexed Belief and Pragmatism

My general methodological approach has been to look for the kind of state that best fits with the practical role of belief ascription, and to identify such a state with belief.<sup>19</sup> I have argued that the practical role of belief ascription is to pick out states that guide action in a systematic manner and that can, in the right circumstances lead to success – this allows us to predict and explain behaviour in an efficient manner, and also intervene to help subjects avoid failure where necessary (through argument or presentation of evidence).

Most important for current purposes is what pragmatism has to tell us about the functional role of belief. First, we know it must be systematic, and not too complex if we are to be able to use it in everyday situations, as opposed to in the laboratory. Moreover, the kinds of features that individuate beliefs must be things we are able to detect with relative ease. That is, dispositions to cause overt behaviour, and dispositions to be altered by environmental states of affairs, rather than subtle differences in, e.g., reaction time or cognitive architecture that are of significance to psychology, but cannot be drawn upon in everyday situations.

There is a clear practical advantage to being able to distinguish inverted subjects. We may well have an interest in knowing whether we are dealing with an expert tennis player or an expert tennis commentator, for example. If we are interested in how someone will perform as a tennis partner, we need to know about their beliefs relative to practical tasks, while if we are interested in what kind of sports column they will write we want to know about beliefs relative to linguistic tasks.

This suggests that *Indexed Belief* has a clear practical advantage over its rivals – as long as the extra predictive power is not bought at the cost of reduced efficiency. In the next section I flesh out the notion of a task to address this potential issue.

## 5. Developing Indexed Belief

### 5.1 Specifying the tasks

---

<sup>19</sup> It is also advocated by many of the philosophers working on the issue of conflicting behaviour. See the final section of chapter 2, for examples of this.

To see whether *Indexed Belief* allows us to predict behaviour *systematically* and *efficiently*, we must look more carefully at the nature of the tasks belief is indexed to. Recall that a task is a category of behaviour; and that a subject has a belief relative to a given task if she has a state which tends to guide behaviour falling within the relevant category, in a manner appropriate for a belief in the relevant proposition. The key issue is what range of behaviour is covered by a given task. It can't be too narrow or too wide.

If it's too narrow, then task-indexed belief attributions will fail to be informative. To see this, suppose we take a maximally narrow reading on which all that's covered is the 'task at hand'. For example, suppose we are looking at what Maya believes relative to the task of getting to La Porte Noire (a bar in Brussels). She heads west from Central Station to get there. So, it would seem, she believes that La Porte Noire is west of Central Station, relative to the task of getting to La Porte Noire (from Central Station on July 1<sup>st</sup> 2014).

The problem is that belief attributions indexed to such narrow tasks give us very little information. The example under discussion only covers her behaviour when she is engaged in an attempt to get to a particular bar (at a particular time, from a particular starting point...). It would not allow us to predict that she would head east from the bar, when aiming for the station. This makes such finely individuated tasks little improvement on simply enumerating a subject's behavioural dispositions. Thus, such attributions would not allow us to predict behaviour systematically and efficiently. The predictive (and descriptive) power of belief attributions only comes out when they apply to a range of behaviour.<sup>20</sup>

However, we do not want to go too wide in the range of behaviour covered, otherwise we will run back into the problem of inversions. When considering Maya's attempt to get to the bar, we do not want to include any of her verbal behaviour. She will not be able to say that it's to the west of the station (we may stipulate), so including this in the range of the task would lead to it coming out indeterminate what she believed about the bar's location (relative to the relevant index). More generally, the range can't include behaviour from both sides of a conflict (at least, for any case of conflicting behaviour we have a practical need to model).

---

<sup>20</sup> This point has long been emphasized by Dennett. See, e.g., his (1971).

The challenge is to specify a stable middle ground – and I think the examples show how it can be met. Maya has a body of fairly coherent behavioural dispositions when it comes to navigating Brussels, guided by her mental image of the city. She is also disposed to make a bunch of assertions about the layout of Brussels that are generally coherent – these are guided by another set of representations. Clearly the two sets of information cannot be integrated with each other, otherwise the behaviour they produce wouldn't conflict.<sup>21</sup> This suggests that the range of behaviour included under a navigational task will be everything guided by her integrated mental image of the city. And the range of behaviour included under a verbal task will include all behaviour guided by the network of representations underlying her assertions about Brussels.

Generalizing, then, if we don't want to run afoul of the problem of inversions, the range of behaviour must all be caused by an *integrated* network of representations – it is taking behaviour from non-integrated states that leads to conflicting results. In order to maximize informativeness, though, it should include *all* behaviour caused by such a network.

It's tricky to articulate the precise borders of such ranges of behaviour. It's going to be an empirical question, determined by our functional structure. However, it's plausible to think we have a pretty good grip on the rough outline of the matter in everyday situations. We can predict that when a sports journalist draws upon a piece of information in writing an article, she will be able to employ it in general conversation about sport, but not necessarily when playing. Similarly, we predict that when a sportsperson draws on information in one play she will be able to use it in others, but not necessarily to articulate it. Therefore, we are well poised to communicate effectively by means of belief ascriptions indexed to such tasks.

## 5.2 Defending Integration

My proposal rests on the assumption that our beliefs divide up into clusters of integrated states – each one guiding a certain range of behaviour in a fairly coherent fashion. This proposal must be defended. First, it should be observed that the default psychological

---

<sup>21</sup> Or rather, if they could be integrated, any conflict could easily be resolved.

picture in philosophy is of a rational agent with a single integrated set of beliefs. However, the examples we've been discussing show this can't be correct. A picture on which there are multiple clusters of integrated beliefs, makes a minimal adjustment to the initial view while accommodating the examples – which gives it *prima facie* appeal.

It's important to be clear that for my view to work, it's only necessary for beliefs to form integrated clusters from a *practical* perspective. My proposal may sound similar to the empirical thesis of *massive modularity*, defended by Carruthers (2006) and Pinker (1997) – which we've seen is highly contentious. The massive modularity thesis claims that beliefs divide up into separate modules, in some psychologically loaded fashion.<sup>22</sup> My view only concerns the functional role of beliefs, as it relates to our behavioural dispositions.

To see this, consider a simplified psychological picture on which we have a single body of beliefs stored in our long-term memory. These beliefs are updated when they are brought into our working memory via various look-up algorithms, and we are able to draw inferences from them.<sup>23</sup> This picture is incompatible with massive modularity since the beliefs in the long-term memory do not divide up into modules in a psychologically significant sense.

However, it is not incompatible with my view. For all I've said, it might be that there are certain groups of beliefs that are all prone to be recalled to the working memory together – due to the nature of our look-up algorithms – but not to be recalled at the same time as those in other groups. For example, there might be certain beliefs that are likely to be looked up in comfortable academic settings, and others that are likely to be looked up in dangerous settings – thus the two groups are, in my sense, inferentially isolated.

My view only requires that our belief states guide coherent ranges of behaviour in an appropriate fashion (i.e. in accordance with the functional role of belief). It also allows for vagueness and a certain amount of overlap in how the clusters are delineated.

One could object that we lack even this level of coherence. Perhaps, we just have a huge number of atomic representational states, each with an idiosyncratic causal role.

---

<sup>22</sup> Exactly what the criteria are for modules varies between theorists, but in all cases it's meant to pick out a psychological kind.

<sup>23</sup> The discussion of psychologism in chapter 1 shows why, though this might be true in part, a model like this very unlikely to provide a complete account of our psychology.

Every state guides some subset of our behaviour, and interacts with some subset of our other beliefs – but this does not happen in a systematic way, and thus doesn't lead to integrated networks.

If such a scenario were the case, it would spell trouble for my view – however, I don't think it's at all plausible. Our behaviour is not so chaotic! If there weren't some systematic structure to the functional role of belief states, it's hard to see how we would be able to attribute beliefs or make predictions on the basis of such attributions with any kind of success. As was mentioned above, we are able to predict pretty well the different ranges of behaviour a belief will guide, for example, when it comes to navigation, talking theory, or playing a sport.

It could be conceded that our belief states form some sort of systematic structure that we can grasp in everyday situations – just not the one I'm suggesting. However, for this to have any force, the burden of proof is on the objector to specify such a structure – otherwise this objection amounts to saying 'well maybe there's something you haven't thought of!' I'm not convinced that such an alternative is forthcoming and so this argument does not seem particularly worrying – I think an inference to the best explanation supports my account.

## **6. Further Issues**

That belief is a three-place relation is a somewhat radical claim, which has significant philosophical repercussions. Before finishing this chapter, I want to look at two that are particularly salient.

### **6.1 Belief Ascription**

I have argued that *Indexed Belief* is the practically significant relation – the one we are led to by the pragmatist methodology. However, it is a three-place relation, whereas our ordinary belief ascriptions seem to express a two-place relation (they make no mention of tasks). There is a question, then, as to how language and state are related. Two approaches are possible here: revisionary and conciliatory. The revisionary approach is to

say that we should scrap our binary belief talk and start using indexed belief instead: this is the regimentation of our natural language practice that best serves our practical goals, and thus, pragmatism seems to favour it.

The conciliatory approach, by contrast, is to rehabilitate binary belief *ascriptions* within the framework of indexed belief *states*. There are a number of ways to make a sentence containing a two-place predicate pick out a proposition involving a three-place property. Presumably if this strategy is workable then it's preferable to revisionism – since it avoids unnecessary overhaul of our current practice, while preserving the practical benefits of the revisionary strategy.<sup>24</sup> The most obvious option is to say that 'S believes that p' is true if and only if *there exists some task t* such that S believes that p relative to *t*. However, this would basically leave us with the contradictory belief view, which has all the problems we've previously discussed.

A more promising approach is *contextualism*. On this view, 'S believes that p' is true in a given context if and only if, for the *contextually relevant* task *t*, S believes that p relative to *t*. This proposal assumes that conversational contexts supply appropriate task parameters. It will allow us to use belief attributions to communicate information about the practically relevant belief states and associated behavioural dispositions at any given time (since these will be picked out by the contextually salient task) – and thus seems well motivated by pragmatism. For example, in a context in which we are wondering whether Maya will be able to make her way to the bar (so that navigational tasks are salient) it will be correct to say 'Maya believes that the bar is west of central station'. On the other hand, in a context in which Maya is giving directions to our friend who we want to meet us (so that verbal tasks are salient) such an assertion would be false.

However, many people find contextualist accounts of this sort implausible, as is familiar from the debate on knowledge ascriptions. The central objection is that such views are allegedly unable to account for cases of inter-conversational disagreement (instances where I judge that someone in a different context spoke falsely when making a

---

<sup>24</sup> More radically, if the pragmatist account of reference I sketched in chapter 1 is correct, the fact that this rehabilitation of belief ascription both fits reasonably well with our current practice and best serves our practical goals gives us reason to think that it is the best interpretation of belief ascription as it now stands.

belief/knowledge attribution).<sup>25</sup> Indeed, this seems a practical deficiency since such disagreement may be useful, so the objection is reinforced by pragmatism.

Fortunately, as is also familiar from the knowledge debate, there are alternatives to contextualism which retain many of the appealing features while dealing with the technical problems. The key point is that we want to tailor what types of task we concern ourselves with to the practically significant features of particular cases. Contextualism is perhaps the most obvious way of doing this: what types of task we use depends on what is relevant in a particular context. However, it is not the only way: competing views such as relativism, expressivism, and subject-sensitive invariantism all offer alternative means of allowing us to be flexible in what we use a term to pick out.<sup>26</sup> Thus, however the wider debates in philosophy of language play out, it would seem possible to combine a practice of binary belief ascription with an underlying structure of indexed belief in a pragmatically acceptable manner.

Whether the revisionary or conciliatory approach is correct depends on whether such a contextualist (or relativist or expressivist or interest-sensitive invariantist) interpretation of natural language belief attributions is plausible. Answering this question requires looking at how the folk use belief ascriptions across a range of contexts. This seems a somewhat orthogonal matter to the issues that have been discussed in this paper so I will not attempt to settle it here.

## **6.2 Modes of Presentation**

The move of indexing beliefs to tasks is similar in form to the well-known proposal to relativize belief to modes of presentation. It might be thought that saying that a subject believes a proposition relative to one index but not another, comes to the same thing as saying they believe the proposition under one mode of presentation but not another. If this were so, my proposal would be open to the charge of reinventing the wheel. I'll explain why I don't think this is the case.

---

<sup>25</sup> See, e.g., Stanley (2005).

<sup>26</sup> See MacFarlane (2005) for relativism, Chrisman (2007) for expressivism, and Stanley (2005) for interest-sensitive invariantism.

There is a sense in which the two concepts are related. Modes of presentation have been invoked to make sense of what are sometimes known as ‘Frege Cases’. As I will be using the term, a *Frege Case* is a situation in which it seems that a subject S believes that  $p$  and believes that not- $q$ , and S appears *rational*, despite the fact that  $p$  and  $q$ , understood as Russellian propositions, appear to be identical – they cannot be distinguished in terms of their truth conditions, or the objects and properties they involve.<sup>27</sup> A canonical example is that Lois Lane believes that Superman can fly but believes that Clark Kent cannot fly, and appears rational in doing so. However, since Superman is Clark Kent, it seems *that Superman can fly* is the same proposition as *that Clark Kent can fly*.

Cases of conflicting behaviour look to fit the description of a Frege case too. For example, Maya believes that La Porte Noire is west of the station in virtue of her mental image that guides her navigation. However, in virtue of her linguistic behaviour, she believes that La Porte Noire is not west of the station. Further, she seems to be rational – we do not generally require people to integrate mental imagery with their linguistic assertions in order to judge them rational.<sup>28</sup>

It might be thought that since the examples of Maya and Lois Lane are both Frege Cases, they both must be explained using the same concept: either both involve belief under different modes of presentation, or neither do (and the same goes, *mutatis mutandis*, for indexing to tasks). I think, though, that the correct lesson to draw from this is that there are two kinds of Frege case.

Cases of conflicting behaviour are importantly different from the traditional Frege Cases that modes of presentation are a response to. Traditional cases, but not my cases, involve pairs of beliefs that are inferentially integrated and guide the same tasks. For example, Lois Lane can use her ‘Clark-Kent-beliefs’ and ‘Superman-beliefs’ in conjunction, to draw inferences: she might believe that Clark Kent will be at the party and that Superman will be at the party, and so infer that she will have two friends at the

---

<sup>27</sup> I’m not claiming that the proper objects of belief are Russellian propositions, rather than the more coarse-grained truth-conditional propositions favoured by, e.g., Stalnaker (1987). My intention is just to avoid ruling out the possibility.

<sup>28</sup> I think we should work with a lenient notion of rationality when defining Frege cases – neither Maya nor Lois’ beliefs resemble the canonically irrational beliefs of someone who believes a false mathematical statement due to bad mathematical reasoning.

party. Moreover, both of these beliefs might guide her in a single task – together, they lead her to go to the party.

An additional feature of note about these cases is that the inconsistency of belief can be resolved *in a flash*, if the appropriate information is provided. As soon as Superman's identity is revealed to Lois Lane, she is able to ascribe a single body of properties to the man, bringing together everything she already knew under either mode of presentation. Note moreover, that these features carry over to the majority of canonical Frege Cases: e.g. Hesperus and Phosphorus, Puzzling Pierre, and John Perry spilling sugar in the Supermarket (i.e. 'the essential indexical').

This is not the case with the main examples in this paper: one of my central claims is that the conflicting beliefs in cases of conflicting behaviour belong to inferentially isolated belief clusters. Moreover, it is clear that in these cases, no single piece of information can be learned to resolve the conflict in a flash. There's no quick fix to make how an athlete plays line up with the sports theory they have read – it requires a lengthy program of training. The same goes for putting one's mental image of a city into words and, unfortunately, for eliminating implicit bias.

For ease of expression, I will from here on refer to the clusters of integrated beliefs as 'compartments'. So in traditional Frege Cases, the conflicting beliefs are in the same compartment – call them *intra*-compartmental Frege Cases – while in the cases I've been focussing on, the conflicting beliefs are in different compartments – call them *inter*-compartmental Frege Cases.

One might wonder what the significance of this distinction is – why *intra*- and *inter*-compartmental Frege Cases are different in *kind*. Crucially, I think, once the *inter*-compartmental cases have been diagnosed, there is no philosophical puzzle remaining. If we accept that our system of beliefs is compartmentalized (as I've argued it is), it's clear how a, loosely speaking, rational agent can have different compartments with inconsistent contents guiding different tasks. Task-indexed belief describes these cases satisfactorily at the doxastic level – though of course there's much of psychological interest that still needs to be explained.

On the other hand, *intra*-modular cases are incredibly puzzling. We still need an explanation of how an agent can use a (Russellian) proposition and its negation in a

single line of reasoning and still appear rational. We need to see how such a case is different from one of straightforward logical error.

Modes of presentation are the tools that are meant to resolve this puzzle. In general, they aim to individuate beliefs more finely by individuating *contents* more finely. This is not just a terminological issue over what is meant by ‘content’; after all, I could define a proposition as being an ordered pair of, say, a Russellian proposition and a task, to create a two-place account of belief, without affecting the substance of the view. The key issue is that modes of presentation are (or at least purport to be) *representational*. Lois Lane, for example, seems to have a complex mental representation of the world with Superman and Clark Kent as distinct parts. Frege, who came up with the idea, certainly saw them as somehow representational: they are the ‘thoughts’ grasped by the mind, and the bearers of truth-value.<sup>29</sup> This is consistent with more contemporary accounts: from Chalmers’ two-dimensional semantics to Peacocke’s inferential role semantics.<sup>30</sup>

On the other hand, the elements of my indexes – tasks – are not remotely *representational* – they simply pick out a category of action. Further, it would not be possible to apply task-indexing to intra-compartmental cases, even if one wanted to. One could not specify the separate tasks to which the ‘Superman beliefs’ and ‘Clark Kent beliefs’ are to be indexed – the states are inextricably intertwined in their functional role, so attempting to specify independent ranges of behaviour for them to guide looks hopeless.

Despite these differences, *pragmatism* is still relevant to the question of traditional intra-compartmental Frege cases. The mental states of subjects in such scenarios influence behaviour in practically significant ways. We could imagine an inversion of Lois Lane who believed that Clark Kent could fly but Superman couldn’t – such a person would make all sorts of outlandish claims. It would be useful to have a notion of belief that could distinguish such subjects, *if* it still allowed for efficient prediction of behaviour. That’s a big if, since existing Fregean theories face a litany of objections, purporting to show they cannot account for the necessary examples in a systematic manner.<sup>31</sup>

---

<sup>29</sup> See Frege (1984).

<sup>30</sup> See Chalmers (2006); Peacocke (1992).

<sup>31</sup> See, e.g., Soames (2002)

Though I do not wish to take a stand on whether this can be done, I think the pragmatic approach allows for a step forward on the issue. Traditional debates between Fregean theorists and their Millian rivals have centred on who can best model intuitive judgements about various belief ascriptions in ever more elaborate scenarios. This is not, to my mind, a promising research strategy. In contrast, the pragmatist project is to account for the *practically significant* features of subjects in Frege cases in an efficient manner – whether this can be done seems like an interesting and tractable question. This shows that even though indexing belief to tasks is not a miracle fix to all puzzles about our mental states, it does fit neatly within an integrated *pragmatist* approach to theorizing about the mind.

### 6.2.1 Exceptional Cases

Though most traditional Frege Cases are intra-compartmental, it's worth looking at whether there are *any* that are in fact inter-compartmental. It would be significant, and welcome, if so – as it would mean such cases could be explained using task-indexed belief, side-stepping the deadlocked debate on modes of presentation.

A clear-cut case of this, if my general picture is correct, is with so-called 'practical modes of presentation'. These are invoked by Stanley and Williamson (2001) as they defend their version of intellectualism. They note that though know-how is a kind of know-that on their view, they still must have some explanation of how and why canonical cases of know-how are so different from canonical cases of know-that.

They argue that when one knows how to do something, one knows the relevant proposition under a 'practical mode of presentation', whereas in typical cases of know-that, one knows the proposition under a 'theoretical mode of presentation'. They attempt to make sense of practical modes of presentation by comparing them to the first-personal mode of presentation that is thought to be involved in cases of indexical belief – and claiming the two are closely analogous.

I think, however, that this analogy is misguided: conflicts involving indexical belief (such as Perry in the supermarket) are intra-compartmental, while the difference between know-how and (typical) know-that goes across compartments. Know-how is knowledge

indexed to practical tasks while typical know-that is knowledge indexed to theoretical tasks. The expression '*practical mode of presentation*' is a misnomer.

A second possible example of this mix-up concerns the super-scientist Mary, introduced in the knowledge argument. The claims I make in order to view the case this way are highly controversial; however, given, the significant philosophical interest in identifying the correct response to knowledge argument, I will provide a sketch of how such an argument could go. To start, here's a version of the problem, adapted from Jackson's (1982) original paper:

*Mary*: Mary is kept in a black and white room and is taught all the neuro- and colour-science necessary to understand colour perception. At time t1 she is told that at time t2 she will be released and shown a red tomato. She says to herself 'at t2 I will be in brain-state X' – where X is the state one is in when one has the experience of seeing a red tomato. At t2, she is duly shown the tomato and infers 'now I am in brain state X'. But she also says 'wow, I'm having an experience like *this* – this is what it's like to have a red visual experience'.

The issue with this, as is probably familiar, is that Mary seems to learn a proposition at t2, but if we're physicalists, she must already know the relevant proposition at t1 – she knows all the physical facts with regard to perception, which constitutes everything there is to know.

There are a huge range of responses to this problem, however one popular response draws on modes of presentation.<sup>32</sup> The idea is that there is a single Russellian proposition *that Mary is in state X at t2*, which Mary knows under a theoretical mode of presentation at t1 and learns under a 'phenomenal mode of presentation' at t2.

An instructive argument against this approach is put forward by Lewis (1988). He asks us to suppose that at t2, instead of seeing a red tomato, Mary learns the Russian translation of the sentence 'I am in brain state X'. At least on some views, she now knows that she is in brain state X under a new mode of presentation. However, this is not

---

<sup>32</sup> See the papers in Alter & Walter (2006) for a range of takes on the strategy.

a significant cognitive achievement – Mary seems to learn something much more substantial in the original case.

This presents us with a challenge: to explain how relearning an old proposition under a new mode of presentation can be a significant cognitive achievement. For this to be satisfactory, one needs to provide an account of the nature of these different modes of presentation that sheds light on Mary's situation. Those who postulate phenomenal modes of presentation believe one can offer an account of them that explains this cognitive significance – for example Loar (1990) argues that they are a special kind of *demonstrative* mode of presentation. However, there has been much scepticism as to whether this kind of response is satisfactory.<sup>33</sup> I want to suggest, therefore, a different approach.

My central claim is that the knowledge Mary gains at t2 is in a different compartment from the theoretical beliefs she possesses at t1 – and so they are indexed to different tasks. To get to this conclusion requires two key premises, which I'll now go through. The first premise is that our everyday beliefs about the physical world are more-or-less insulated from our esoteric scientific beliefs. The everyday beliefs roughly correspond to those we express in folk vocabulary and the esoteric beliefs correspond to those expressed in theoretical vocabulary and in scientific inquiry. As motivation for this, consider the following example:

*Sanaya the engineer:* Sanaya is inside the Stata Centre planning on knocking through some walls on the 8<sup>th</sup> floor, in order to turn the library into a shrine to Bob and Judy. She has seen the various parts of the building and is disposed to navigate it effectively, avoiding objects, walking to the rooms she needs to get to, reaching for door-handles, light switches etc. without looking. This action is guided by the principles of 'folk physics'. Sanaya also has memorised information found in a blueprint of the building and more general principles she learned in her engineering PhD. She uses this when drawing up a plan for how to build the doors, telling her foreman what to do etc.

---

<sup>33</sup> See e.g. Ball (2009)

The two groups of information are not consistent: in particular, some elements of folk physics are incorrect, and are contradicted by the principles of engineering. Further, each set of beliefs can only be drawn upon in some of the relevant contexts. Sanaya can't draw upon her knowledge of the blueprint when moving sofas about, and she can't use her practical knowledge to fill in the gaps if she missed a bit of the blueprint. Further, the two groups of information are not easily integrated: note that it is not realistically possible to modify the folk physics one relies upon in everyday tasks to accord with high-level engineering theory. This suggests that Sayana has beliefs belonging to two different compartments, indexed to everyday tasks and theoretical tasks respectively.

The second premise is that we can extend this idea to our own minds – another kind of physical entity. We have theoretical beliefs about the brain that play a role in discussing neuroscience and other theoretical activity. We also have everyday beliefs about minds that are expressed in experiential language (I'll refer to these as *phenomenal beliefs*). Again the beliefs that guide us when doing neuroscience are not easily integrated with our phenomenal beliefs – the ones we use when describing our own experiences, and those of others. The phenomenal beliefs guide practical tasks: we can imagine mental states and recall experiences in order to predict our own behaviour, to predict the behaviour of others, and to guide our understanding to what happened to someone else. Moreover, these beliefs are directly updated by perceptual experience.

Therefore, if we accept the second premise, we can conclude that everyday experiential beliefs and neuroscientific beliefs belong to different compartments, and are indexed to different tasks. The above discussion was not intended as a watertight argument for these two premises, but just to show how they fit naturally with the overall view I have been arguing for. Now I'll return to the knowledge argument.

On my view, at t1 Mary knows that she is in brain state X relative to neuroscientific tasks, but not relative to everyday tasks. At t2, upon seeing the tomato, she gains the belief that she is in brain state X relative to everyday tasks – the information is now contained in a new compartment. This grants her all sorts of new abilities – she has

additional ways to predict her future own mental states and behaviour, as well as those of other people.<sup>34</sup>

We can see that this is a significant cognitive achievement – it transforms her ability to think and act in everyday situations in important ways, and so it meets Lewis’ challenge. Further, since Mary gains a belief in a new compartment rather than an additional belief in the compartment of her original theoretical belief, talk of her having a belief under a new ‘phenomenal’ mode of presentation is ill-conceived.

This is not a complete refutation of the knowledge argument. I make no claims that acquiring a belief relative to a new task gets to the bottom of the peculiar and mysterious nature of conscious experience. I do, though, think that the tools of indexed belief-allow us to see one aspect of Mary’s cognitive development, and to diagnose why appeals to phenomenal modes of presentation are dubious.

## 7. Conclusion

Over the course of this chapter, I have discussed the complexity of cases of conflicting behaviour. There is much more to the issue than simply conflict between elite and non-elite states – and thus the problem is much more complex than might initially have been supposed. I have argued that this compels us to adopt a somewhat radical solution: *Indexed Belief*. Finally I have shown some ways in which this thesis connects to some well-known topics in philosophy of mind.

---

<sup>34</sup> Interestingly, I think this allows us to see what Lewis (1988) got right and what he got wrong in his response to the knowledge argument. He claimed that Mary acquired know-how but not know-that. On my view, acquiring know-how where one previously only had (typical) know-that, amounts to acquiring knowledge relative to a practical task that one previously held relative to a linguistic task. This is similar to what I say occurs with Mary – she too acquires belief relative to a new task. However, not all cases of acquiring beliefs relative to new tasks are cases of know-how acquisition. Because my view gives us the power to distinguish many kinds of task – not just practical and theoretical – more options are available. The everyday tasks that phenomenal beliefs are indexed to are different from the practical action-based tasks (typical) know-how is indexed to. Describing Mary as acquiring know-how always seemed awkward: it’s hard to say *what* she learned how to do – it didn’t seem like any form of skilled action. The more nuanced task indexed account allows us to more carefully model the case.

[blank by design]

## Chapter 5

# Epistemic Ideals

### 1. Introduction

In earlier chapters, I argued that the concept of rationality is closely linked to the practical role of knowledge ascription – and in particular its relationship to the *normative* significance of knowledge. In this chapter, I will further explore this issue by looking at the notion of an *epistemic ideal*.

Much work in epistemology has focused upon idealized rational agents. Bayesian approaches to rationality, in particular, look at formal constraints upon sets of credences that are meant to capture what it is to be perfectly rational. Since imperfect creatures like us seem to have little chance of complying with such an ideal, a natural concern is what consequences it has for us. I think this question underlies a number of philosophical problems related to Bayesian approaches to epistemology.

In this chapter I will look at three puzzles that arise in connection with the Bayesian ideal: the question of why the Bayesian ideal has greater significance than complete omniscience; Chalmers' (2011) apparently compelling argument that the Bayesian ideal commits us to a substantial body of a priori knowledge; and the tension between theoretical considerations and intuitive judgement when assessing 'level-bridging principles' regarding higher-order evidence.

Next I will investigate the nature of ideals in general to shed some light on these puzzles. I first distinguish two kinds of ideals: pragmatic Galilean ideals and robust Aristotelian ideals. Galilean ideals have two key features: first, there may be multiple legitimate idealizations of a single system; and second there may be 'artefacts of the system' – that is, features of the model that do not tell us anything significant about reality. The principal aim of this chapter is to show that epistemic ideals are Galilean, and

explore the consequences for epistemology.

The first consequence is that there are multiple epistemic ideals – in particular, alongside the Bayesian ideal, there is also what I call *idealized epistemic resource management*, which is more closely tied to our cognitive architecture. I will motivate the claim that this ideal is worthy of serious philosophical investigation.

The second consequence is that we gain fresh traction on the aforementioned epistemic puzzles: the *practical* importance of Bayesianism shows why it is more significant than an ideal of complete omniscience; the potential for artefacts of the system allows us to resist Chalmers' argument; finally, we can hold on to theoretically appealing level-bridging principles without having to reject common-sense judgements about particular cases.

## **2. The Puzzles**

I will now describe three puzzles related to Bayesian ideals, and suggest that we are not well-placed to make sense of them if we do not clarify the background theory. This motivates looking more carefully at the nature of an epistemic ideal. In what follows I will argue that we have the tools to make progress on these puzzles if we understand the Bayesian ideal as Galilean – as I claim it is.

### **2.1 The Happy Medium Puzzle**

Obviously Bayesianism does not provide a picture of our actual epistemic state – most pertinently it requires logical omniscience – that's why it's an *ideal*. A familiar worry, though, is why it's the Bayesian model captures that the optimal epistemic state rather than complete omniscience. This point is made by Foley, who puts it as follows:

If a logically omniscience perspective... is an ideal perspective, one to which we aspire and one that we can do a better or worse job of approximating, so too is an empirically omniscient perspective. If this were a reason to regard all departures from logical omniscience as departures from ideal rationality, it would be an

equally good reason to regard all departures from empirical omniscience as departures from ideal rationality. But of course no one wants to assert this. (Foley 1993: 161)

Instead of seeing this as an argument against the significance of a Bayesian ideal (as I think Foley wants to do), I think we should regard it as a *puzzle*. As many defenders of Bayesianism have noted, there is clearly a notion of ‘epistemic ideal’ that doesn't fit with complete omniscience. Namely, a *normative ideal* – one it is useful to *hold us to* (here, and throughout, I am understanding ‘normative’ in an evaluative sense). It makes no sense to criticize someone for not knowing an empirical fact about distant entities they've had no contact with – this is not something we can aspire to do. In this vein Greaves and Wallace (2006) say “Real epistemic agents are not (at least not quite) like this [i.e., like the Bayesian ideal]. Bayesian epistemology is a normative theory rather than a purely descriptive one” (608). Similarly, Schoenfield (2015) claims: “We are only interested in procedures that are, in some sense, available to the agent upon undergoing the learning-experience... [we] are interested in idealized information processing, and not idealized information possession.”

It's highly plausible that there is a normatively significant epistemic ideal, and that it must fall somewhere in between our actual psychological structure (replete with all our quirks and errors) and complete omniscience. We have a sense of ‘good epistemic practice’ and take epistemic praise and criticism seriously in everyday life.<sup>1</sup> Further, chapter 2 showed that epistemic success is required to further mutual goals in a cooperative society. This demonstrates the need for a normative ideal that models good epistemic practice, so we can praise people for accordance with the ideal and criticize them for deviation. Granting this provides two constraints: on the one hand we want an ideal that we can reasonably aspire to – one that, as Schoenfield puts, it involves ‘procedures available to the agent’, rather than clairvoyant knowledge of distant objects; on the other, the ideal must correct our mistakes, in order to show us where we've gone wrong.

---

<sup>1</sup> Usually, this would be done colloquially rather than in epistemological language: people appreciate being called ‘smart’ and resent being called an ‘idiot’. These terms *can* (though not always) be used to attribute rationality and irrationality respectively.

Clearly, the Bayesian ideal does fall somewhere between our actual psychological structure and complete omniscience. However, there are many other middle points, since we could idealize away from our psychological structure along all sorts of dimensions in ways that stop short of omniscience. Consider, for example, a ‘perceptual ideal’ on which an agent has perfectly reliable perceptual faculties, immune from illusion and hallucination, but who is capable of logical errors. This gives us the *Happy Medium Puzzle*: Why should we think Bayesianism provides the normatively significant midpoint between our actual psychological structure and omniscience?

There has been recent discussion that pertains to this issue but, though insightful, I don't find it fully satisfactory. I'll look at some recent arguments in defence of Bayesianism and argue that they don't fully solve the happy medium puzzle – suggesting that it's worth looking more carefully at the nature of epistemic ideals, in order to find a solution.

One thing that speaks in favour of Bayesianism is its utility. It provides a model that has fruitful applications in a range of places. First, it provides a model of agents' epistemic states that is of use in the social sciences – in particular, classical economics has produced a range of fruitful predictions using the Bayesian ideal to model agents' behaviour. Second, in many places, reasoning according to Bayesian norms has proved incredibly useful in particular circumstances – for example in gambling and sports analytics.<sup>2</sup> Finally, the Bayesian model has proven incredibly useful within philosophy itself – from philosophy of science to formal epistemology. This all suggests that Bayesianism must be philosophically significant; however, it doesn't provide an explanation as to *why*, and does not provide evidence that it is a *normatively* significant middle ground.

Further, there is a plausible-seeming argument against the normative significance of Bayesianism that needs answering. The argument stems from the well-known principle that *ought implies can*. It's tempting to think that if something is a normative requirement it must be possible, in some sense, for us to accord with it – this is why empirical

---

<sup>2</sup> What's useful is not just probabilistic reasoning, but Bayesian rules of inference in particular. There was dispute over whether such scenarios should instead be approached using frequentist rules of inference. The Bayesian method turned out to be far more effective. See, e.g., Hájek (2003) for the contrast between frequentism and Bayesianism.

omniscience is an unacceptable norm. However, the Bayesian ideal requires, among other things, complete probabilistic coherence amongst all our beliefs – something we have no hope of achieving.

Bayesian enthusiasts are quick to point out that the reading of ‘ought implies can’ being appealed to may not apply to epistemic normativity. Smithies (2015), draws on Feldman (2000) theory of ‘role-oughts’, to make this point. Feldman explains the notion as follows:

There are oughts that result from one’s playing a certain role or having a certain position. Teachers ought to explain things clearly. Parents ought to take care of their kids. Cyclists ought to move in various ways. Incompetent teachers, incapable parents, and untrained cyclists may be unable to do what they ought to do. Similarly, I’d say, forming beliefs is something people do. That is, we form beliefs in response to our experiences in the world. Anyone engaged in this activity ought to do it right. (Feldman, 2000, 676)

So on this picture, we have a thin obligation to accord with the epistemic ideal simply in virtue of occupying the role of a believer – but this is not the substantive kind of obligation that upholds an ought-implies-can principle.<sup>3</sup> I think this argument is sufficient to show that our inability to fully realize the Bayesian ideal does not *necessarily* undermine its normative significance. However it emphasizes the aforementioned puzzle, as to why these norms apply to *us*. To solve this one first needs to show why the Bayesian ideal best captures the ‘role of the believer’. Second, one needs to explain why the ‘role of believer’ captures a significant notion of epistemic normatively. What determines the role? If it’s simply a matter of conceptual analysis regarding our concept of belief, it’s not clear why we should aspire to accordance with this role over a deviant role which allows, say, for certain inconsistent beliefs.<sup>4</sup>

A more sophisticated response is offered by Smithies (2015). He notes that though we

---

<sup>3</sup> This line of response is also advocated by Christensen (2004) and Schoenfield (2015).

<sup>4</sup> This point is analogous to Enoch (2006)’s ‘agency/schmagency’ objection to constitutivism about moral normatively. Note, though, that unlike Enoch, I don’t take this to be a decisive objection, but a call for further investigation.

cannot fully realize the Bayesian ideal, we can *approximate* it. In his words: “The ideals themselves need not be humanly achievable so long as we can make sense of better and worse approximation towards those ideals... our ordinary standards of rationality are best understood as approximations towards ideal standards that take our contingent human limitations into account.”<sup>5</sup>

This response leaves questions unanswered. Though there are ways we can get closer to the ideal – for example, by correcting various inconsistent beliefs – we cannot approximate it in the technical sense. That is, we cannot get arbitrarily close to the ideal: it is impossible for us, say, to eliminate all but ten of our inconsistent beliefs. The reason for this is that there are, in virtue of our cognitive structure, hard barriers that stand in the way of us achieving the ideal of perfectly-integrated, probabilistically-coherent credences – as has been discussed at length in earlier chapters. For example: our information states are quite plausibly modular; we have both binary and partial beliefs with distinctive functional roles; we have long-term and working memory; and items in the long-term memory are only updated if they are brought into working memory, so our total set of credences cannot be updated together.<sup>6</sup> These features do not come in degrees: we possess them, the Bayesian ideal does not, and we cannot approximate not possessing them.

We could create a model that built in these features while requiring logical consistency not ruled out by any of the relevant barriers. We also approximate this model in a loose sense, and one could hold that this is better suited to be the normatively significant happy medium. Smithies (2015) contends that the more psychologically loaded ideal is *derivative* from the Bayesian ideal. However, in order to vindicate this claim, one must account explain what makes one idealization more fundamental than another, which requires understanding what accounts for an ideal’s normative significance. This requires a deeper investigation of the nature of ideals (my task over the course of this paper).

## 2.2 The A Priori Puzzle

---

<sup>5</sup> Smithies (2015), p. 17.

<sup>6</sup> I provide further explanation of these phenomena, along with relevant citations in section 5.

The second puzzle concerns Chalmers' recent defence of the a priori. In an influential paper – Chalmers (2011) – he argues that use of the Bayesian ideal commits us to the existence of a substantive body of a priori truths. Thus, he claims that we are faced with a choice: reject the Bayesian picture of rationality or accept the a priori. This seems a surprising conclusion since many a priori sceptics embrace the Bayesian model. Moreover, the Bayesian framework doesn't seem like the sort of thing that *should* impose substantive epistemological theses. The appeal of the framework is that it allows us to model and evaluate epistemic states using probabilistic machinery. It seems like it should be possible to do this while rejecting the a priori.

This gives us what I'm calling *the a priori puzzle*: why does there appear to be a tight connection between Bayesianism and the a priori? To get a better sense of the problem I'll give an outline of Chalmers' argument.

Chalmers' argument builds upon the so-called 'front-loading' response to Quinian objections – which he assumes are the most compelling argument against the a priori.<sup>7</sup> One begins by taking a *prima facie* a priori statement – for example 'all cats are animals'. One then looks to the potential empirical evidence Quinians take to undermine the statement's a priori status – for example, if we learned that all the creatures we call 'cats' are in fact robot spies sent from Mars, we'd conclude that 'all cats are animals' is false. The *front-loading* move is to put this undermining evidence into the antecedent of a conditional – so we have 'if the creatures we call 'cats' are in fact robot spies sent from Mars, then not all cats are animals', and 'if current zoology is roughly correct regarding the creatures we call 'cats', then all cats are animals', which are both allegedly a priori.

The strategy is counter-example-proof, since whatever undermining evidence the Quinian presents can be built into the antecedent. It has been argued by some – e.g. Williamson (2008) – that this process of providing undermining evidence can be continued *indefinitely*. To allow for this, Chalmers introduces the notion of a *maximal evidence statement* – such a sentence specifies all the evidence a subject could possibly receive. For a conditional 'If E then p', where E is a maximal evidence statement, Chalmers contends that if such a statement is known, it is known a priori – it is immune

---

<sup>7</sup> See Hawthorne (2002), White (2006) for earlier elaborations of this argument.

to being undermined by empirical evidence, since all possible evidence is built into the conditional assumption.

A crucial problem with this argument is that it depends on our attitudes towards *sentences*. One issue here is that even the conditionals Chalmers specifies are not immune to revision if the meaning of the words they contain changes. Thus for there to be a priori statements immune from revision in the way Chalmers requires, there must be a non-question begging way to identify changes in meaning.<sup>8</sup> A second issue is that since these sentences involve maximal evidence statements, they must be massively complex (perhaps even infinitely so!). As such, they are not sentences we could possibly comprehend or utter, never mind endorse – thus it's hard to see how they can be a priori for us.

These issues motivate Chalmers to move to the Bayesian framework. Instead of looking for a sentence an agent is inclined to endorse come what may, he suggests we look for a statement that an ideal Bayesian agent has high credence in conditional on any possible evidence. He argues that any such statement will be a priori. Moreover, he suggests this account allows us to solve the problems facing the front-loading argument. Chalmers argues that a sentence will have changed its meaning if and only if the agent changes her credences conditional on maximal evidence statements (thus the Bayesian framework gives us a non-question begging account of meaning change). The rationale here is that if one changes one's attitude towards a sentence without receiving any new evidence, one's understanding of what the sentence means must have changed. The Bayesian framework also allows us to sidestep the complexity issue. It's built into the model that for any proposition  $p$  and evidence  $E$ , the agent will have a well-defined conditional credence  $Cr(p|E)$ .<sup>9</sup>

He notes that if  $p$  is a sentence and  $E$  is a maximal evidence statement, then if an ideal Bayesian agent has high credence that  $p$  is true conditional on  $E$ , then the agent will have high credence in the material conditional ' $E \supset p$ ', conditional on any evidence whatsoever.

There are a couple of initial objections to Chalmers that I want to put to one side.

---

<sup>8</sup> A related point is that the revision of attitude must be *rational*. Chalmers thinks the Bayesian move can also solve this problem.

<sup>9</sup> Note that this also allows us to answer the rationality question, since a Bayesian ideal is by construction rational.

First, one might point out that in order to secure a *substantial* body of a priori statements, Chalmers must rule out subjective Bayesianism. On this view, any probabilistically coherent body of credences is rational, so no non-trivial propositions will be rationally guaranteed to be given high credence conditional on all evidence. Chalmers claims that subjective Bayesianism is overly sceptical about rationality, so need not be taken seriously by the defender of the a priori. I think this is reasonable if we grant that Bayesian ideal captures the normative notion of rationality. Given the role rationality plays in promoting mutual goals, it's implausible to think that subjective Bayesianism can provide the whole story – avoiding probabilistic incoherence would not, on its own, maximize successful action.

One might respond by claiming that we do not have to derive the whole story on rationality from the Bayesian ideal. However, this if the ideal agent does not have to get everything right, it is not clear what being a normative ideal amounts to. Therefore, upholding this line of argument requires better understanding the nature of such ideals. (I think this line of response is essentially correct, but to adequately defend it, one must look in depth at the background theory.)

A second worry concern Chalmers' claim that the priors of the rational ideal must be a priori. One might think that determining the correct priors can be an *empirical* matter. However, on a standard definition of 'a priori', if a proposition is not a priori, one must *acquire evidence* in order to have reason to believe it – or at least, it must be that one could acquire evidence that would require one to reduce one's confidence in it. Chalmers' argument entails that, granting Bayesianism, the ideally rational agent will not revise her credence in the relevant propositions no matter what evidence she receives – which seems to entail that they are a priori. One might claim that the Bayesian ideal does not tell us everything about how *we* should respond to empirical evidence, and thus does not determine what is a priori for us. As with the previous objection, though, justifying this argument requires looking more carefully at the notion of an epistemic ideal.

### **2.3 The Level-Bridging Puzzle**

As has already been mentioned, the notion of evidence plays a crucial role in the

Bayesian model – one's rational credences are meant to be determined by 'what one's evidence supports.' However, the nature of evidence remains a matter of controversy. The issue of higher-order evidence is a part of this: this is the question what one's evidence says about one's evidence. More precisely, suppose your evidence supports proposition  $p$  to degree  $x$ , and let  $E_y$  be the proposition that your evidence supports  $p$  to degree  $y$ . What can be said about the extent to which your evidence supports  $E_y$ , for any  $y$ ? There is strong theoretical motivation for saying that your evidence supports  $E_x$  to degree 1 (and your evidence supports  $E_y$  to degree 0 if  $x \neq y$ ). Or to put it another way, the rational agent is omniscient with regard to what their evidence supports. Constraints that connect first-order and higher-order evidence in such ways are often called 'level-bridging principles'.

The key point here is that the Bayesian model is much more elegant when the relevant higher-order evidence principles obtain. A pertinent example of this is revealed in a recent exchange between Greaves and Wallace (2006) and Schoenfield (2015). Greaves and Wallace argue that, granting certain assumptions, on the best account of the Bayesian ideal one updates by conditionalization. That is, for any proposition  $p$ , if one receives evidence  $E$ , one's new credence  $Cr(p)$  should be equal to one's prior credence  $Cr(p|E)$ . Schoenfield observes, however, that one needs a crucial assumption for the argument to go through: that the rational agent is certain that for every proposition she thinks she might learn, she will learn  $p$  if and only if  $p$  obtains (here 'learning  $p$ ' is equivalent to acquiring  $p$  as total evidence). Such an assumption amounts to a strong access constraint on evidence – i.e. the rational agent must be omniscient with respect to one's evidence. If one rejects this principle, one must change the rational updating rule from conditionalization to the less attractive principle Schoenfield calls 'conditionalization\*', which states that upon learning  $p$ , an ideally rational agent will conditionalize on the proposition *that she learned  $p$* .

Despite these theoretical benefits, there seem compelling arguments against level-bridging principles. In particular, there are a number of examples in which it is intuitive to think that the principles fail. For example, I may have derived a watertight proof of mathematical theorem  $T$  (and so have strong evidence for  $T$ ), but then be told that I have been slipped a reason-distorting drug (so I have evidence that my evidence for  $T$  is

weak). Other philosophers discuss a range of similar examples.<sup>10</sup> If the Bayesian wants to take such examples at face value, she must move to a more complicated model, that uses something like conditionalization\*.

The *Higher-Order Evidence Puzzle* concerns how to weigh the competing theoretical and intuitive considerations. Theoretical considerations strongly recommend level-bridges, while looking at applications seems to recommend rejecting them. I will show how, in this case too, a better understanding of ideals allows us to make progress with the puzzle.

### 3. Two Kinds of Ideal

An ideal is a *model* of an actual entity (or type of entity) – one that holds some features fixed, alters some parameters to idealized values, and ignores some entirely.<sup>11</sup> A familiar example concerns idealized frictionless systems. Here we take a system of physical objects and hold their mass, velocity etc. fixed, while the presence of friction and air resistance are ignored.

It's less familiar to think of something like an idealized moral agent in these terms, but it is possible to describe it as such. In this case we take an actual agent and hold various non-moral features fixed, such as her physical capabilities and empirical knowledge; and we set various moral features to ideal settings – perhaps changing her motivational framework to include a desire to follow various moral rules, and giving her the ability to resist temptation.

To put it more intuitively, when using moral ideals, we generally want to answer the question ‘what would an ideal agent do in this situation?’, then we can assess the actions of an actual agent relative to this. To do this, we look at the actual agent and create a model of her that holds the non-moral features fixed and sets the moral features to the

---

<sup>10</sup> For example, Elga (2013) discusses ‘the puzzle of the unmarked clock’ in which our visual evidence of the position of the clock’s hands is more precise than our evidence about how accurate our appraisal of the visual evidence is. Further, Schoenfield (2015) presents cases in which an agent acquires total evidence E, but cannot be sure that she does not acquire additional evidence as well – i.e. her evidence does not support the claim that E is her total evidence (note that this is a more modest claim than the proposal that one’s higher order evidence can entail that one’s first order evidence is *wrong*).

<sup>11</sup> Strictly speaking, this third feature is just a special instance of the second: one takes a parameter and alters it to the idealized value of zero.

idealized values. Then we look at what an agent who matched the ideal model *exactly* would do in the situation, and compare this action with that of the actual agent.<sup>12</sup>

Similarly, a Bayesian ideal is used to answer the question ‘what would an ideally rational agent believe in this situation’. The Bayesian ideal is a model of an actual agent’s epistemic state that holds the *arrational* features fixed while setting the rational features to idealized values. The arrational features would include the agent’s evidence, and possibly other features such as her priors, depending on the version of Bayesianism. The rational features are the agent’s beliefs or levels of confidence, and again possibly other things such as the agent’s priors – the ideal will replace the actual beliefs with a probabilistically coherent set of credences appropriately related to the evidence and (possible idealized) priors. One can also consider the generic ideal that models the features shared by ideally rational agents across *all* situations without specifying the non-moral features. This generic ideal can then be used to determine the specific ideal for any given situation.

A controversial issue I wish to put aside is the *metaphysical* status of such ideals. There is longstanding disagreement about the nature of models in general, that does not need to be resolved for our purposes. Options defended in the literature include: physical objects, set theoretic structures, and fictional entities.<sup>13</sup> This debate hinges on deep questions about general ontological issues that I do not want to get into here. Crucially, everything I say about epistemic ideals (perhaps with a little paraphrasing) will be compatible with all the competing views. It’s clearly possible to create the ideals with set-theoretical models – mathematical models of the Bayesian ideal have already been created and discussed at length. Moreover, the idealized situations we are concerned with are *possible*, at least if one does not have a very narrow notion of possibility, so these possible physical entities will also work as models – and, granting this, there can certainly be fictional entities matching the ideals.<sup>14</sup>

---

<sup>12</sup> A more elegant account can be provided if we adopt a definition of ‘situation’ on which it includes not just external circumstances of an agent but also a specification of all non-morally relevant features of an agent (e.g. physical dimensions, empirical beliefs or aesthetic preferences). This would allow us to ask directly what an ideal agent would do in a given situation, without making reference to an actual agent. Note though that the relevant non-moral features of an actual agent would then play a role in determining what situation she was in.

<sup>13</sup> See Frigg and Hartmann (2006) for a comprehensive discussion of the possible views.

<sup>14</sup> One crucial feature of ideals I will rely on is their relation to counterfactual claims: I will talk about how

For ease of expression, I will talk about ideals as though they are entities of the same nature as the things they are idealizing. So I will specify an ideal as follows (letting S be an entity and S\* be a model of S): 'S\* is an ideal for S if S\* shares features *abc* with S, and S\* has idealized properties *xyz*.' Note though that this could be easily paraphrased in terms of S\* being a model that *represented* an entity with the appropriate properties, if one preferred working with an alternative theory of models.

For our purposes, the metaphysical status of ideals is much less important than their applications. There are various reasons one might appeal to an ideal. Some of the key ones are as follows:

1. *Prediction*: Using an ideal allows us to make accurate predictions about actual systems.
2. *Explanation*: An ideal can be part of good explanation of actual events. (One might explain what actually happened by describing what would have happened in idealized 'normal conditions' then pointing out how interfering factors produced the actual results.)
3. *Normativity*: An ideal may provide normative force. (Cases where an actual entity deviates from the ideal are in some sense bad, while cases where they accord are good.)
4. *Practicality*: Use of an ideal can promote various practical goals. (I'll get to examples of this later.)

Clearly, these uses are not mutually exclusive; however, which are emphasized can lead to more or less theoretically loaded notions of an ideal. For example, for an ideal to provide good predictions all it has to do is *work* – the reason why it yields useful results doesn't matter. On the other hand, on certain theories of explanation a good explanation must, in some sense, get at the 'joints of reality', so for an ideal to play a role in a robust

---

the ideal agent would have reasoned or acted. I think, though, that whatever theory of models one adopts, one will be able to make sense of this feature. Roughly speaking, one looks to the possible world in which the agent in question is (in the relevant sense) isomorphic to the ideal (whatever form this model takes) and looks at how the agent reasons or acts in such a counterfactual situation.

explanation, it must have some sort of privileged metaphysical status.<sup>15</sup>

I want to start by looking at ideals that are practical in nature. Here we can take frictionless models in mechanics as an example. In the correct circumstances, this model allows us to make fairly accurate predictions about how actual physical bodies will interact while keeping our calculations manageable. In this case it is our predictive goals (and practical constraints on calculation costs) rather than the intrinsic nature of the object that determines which ideal is applicable.

There are two features of such an ideal that are of particular interest. The first is that it creates the possibility of ‘artefacts of the system’ – features of the model that have no bearing on reality, and are merely there to allow application of the model proceed more smoothly. For example, frictionless ideals are deterministic, but this in no way impugns the indeterminism of the actual universe – or suggests that the world in some sense ‘aspires to determinism’.

The second feature is that there may be more than one legitimate idealization of a single entity, relative to different goals. Take as an example, a system of billiard balls set in motion on a table. The frictionless ideal works well for predicting, say, the immediate outcome of a collision between the balls – what direction and speed they will rebound. However, it is very bad at predicting where the balls will come to rest – indeed it will say they will keep on bouncing around for eternity. If we wanted to predict the resting place of the balls, we would need to employ a more complicated Newtonian ideal which didn't ignore friction. Moreover, if we wanted to make incredibly accurate predictions and calculation was no cost, we might have to move to a microphysical model.

Following terminological convention from the discussion of models, I will call such practically motivated ideals *Galilean Ideals* – so-called because of Galileo's famous work on frictionless systems. In contrast to this notion, we also have room for a more robust ideal, which I'll now discuss.

We can get a grip on this robust notion by looking at an idea associated with Aristotle: the ideal an entity is associated with is determined by its *nature*. For example, it is a person's nature to be rational, so an ideally rational person is just a person who

---

<sup>15</sup> This view of explanation is defended by Sider (2011).

fully realizes their nature.<sup>16</sup> There is clearly a privileged connection between a given entity and a given ideal, grounded in the entity's nature. One could use this framework to defend Smithies' claim that the Bayesian ideal was fundamental and the psychological ideal derivative – if one could show that it was in a person's nature to satisfy the Bayesian norms of rationality.

We can draw out some interesting features that seem to go along with this type of ideal. First, there is not room for artefacts of the system: since the ideal is determined by certain components of the objective world, there will not be features of the model purely there to allow for more convenient applications, that the actual entity shouldn't be held to.<sup>17</sup> An entity aspires to fully realize this kind of ideal. A second crucial feature is that there will only be one ideal corresponding to a given entity (at least in relation to a given dimension of idealization) – since things only have one nature.

Though Aristotelian natures are perhaps the most vivid example of a metaphysically robust connection between entity and ideal, they are not the only possibility. One could alternatively appeal to a theory on which there are certain metaphysically privileged properties that ordinary objects only partly instantiate, while the ideal fully instantiates them – this, provided there was a respectable account of 'partial instantiation', would be an update of Plato's theory of forms. Alternatively, a theist might claim that ideal for an entity is determined by how God desires it to be. Finally, one might simply suggest that there are objective normative facts, and these suffice to make the connection. Following the literature on models, I will call all such robust ideals *Aristotelian Ideals*.<sup>18</sup>

It's important to note that these two kinds of ideal do not give a dichotomy of all possible cases, but instead point to extremes on a spectrum. Still I think they provide a useful way of framing our investigation. Simply saying that Bayesianism provides an 'epistemic ideal' is insufficiently precise. We need to determine whether they are more Aristotelian or Galilean in nature. In what follows, I will argue for the latter. More

---

<sup>16</sup>Strictly speaking, all idealization should be indexed to kinds. So the ideal for a given person qua person is to be a rational being. Perhaps the ideal for that same person 'qua teacher' is to always be explaining things clearly.

<sup>17</sup>Note that when working with an Aristotelian ideal, one might modify the model to allow for easier application – and this modification might well bring with it artefacts of the system. However, this would be a *hybrid* ideal containing both Aristotelian and Galilean elements – and the artefacts of the system would involve only the latter.

<sup>18</sup>See Frigg and Hartmann (2006).

precisely, I will argue that insofar as Bayesianism provides a legitimate epistemic ideal, it must be Galilean rather than Aristotelian.

#### 4. Normative Ideals

Galilean ideals most obviously apply when our aims are purely predictive. However, as was discussed above, rationality is a *normative* notion – this was a central premise in the Bayesian enthusiast's defence against the sceptic when responding to the happy medium puzzle. There may be some cases in which applications of Bayesianism are purely predictive; for example, in economics, as was mentioned above. But this is not all we do with epistemic concepts. We explicitly note cases in which people deviate from the ideal, and criticize them as irrational – rather than treating them as conforming to the ideal for the purposes of simplification. One might think this makes the view that epistemic ideals are Galilean misguided – resting on an equivocation of the term ‘ideal’. I take Smithies (2015) to be making such a point:

We can distinguish genuinely normative idealization from the non-normative idealization that is familiar in science. It can be convenient to ignore false predictions of a scientific theory when it sufficiently approximates towards the truth, especially when these are mere side effects of mathematical machinery that is otherwise useful or in-dispensable... they [i.e. epistemic ideals] are often thought to involve non-normative idealizations in much the way as scientific theories... The rival view defended in this paper is that logical omniscience requirements are idealizations in the normative sense that they make true claims about what ideal rationality consists in.

However, I think this dismissal is too hasty. I will show that it's possible for Galilean ideals to do normative work. We saw in chapter 2 that the practical role of certain concepts can concern *promoting* as well as predicting behaviour. This observation can be extended to the practical role of an ideal. I will show how this works by analogy with an expressivist account of our ethical concepts, before returning to the epistemic case in the

next section.

Our starting point is a rather crude appropriation of the ethical theory sketched by Gibbard (1992).<sup>19</sup> The key motivating observation is that in a cooperative society, it will be in everyone's interests to coordinate their behaviour so everyone acts in a manner that promotes mutual benefit – for example, it will be in everyone's interest if no one goes on unprovoked killing sprees. Call these the *cooperative goals*. Establishing a system of praising and blaming people can play an important role in this project – one praises behaviour that promotes the cooperative goals, and criticizes behaviour that undermines them. On the meta-ethical position we're considering, it is the role of ethical concepts to promote the cooperative goals – calling an act 'right' expresses the relevant praise and calling it 'wrong' expresses the relevant blame.

It's possible (though by no means necessary) for an *ideal* to play a role in this practice. An idealized version of an actual agent acts in a way that would promote the cooperative goals in the actual agent's situation. Recall that we might construct a moral ideal by holding fixed the actual agent's beliefs and capacities to act, but changing her desires to include a preference for promoting various mutual goods, and ignoring the presence of temptation, weakness of the will etc. When an agent acts as the ideal agent would act, she is to be praised; when she deviates from such action, she is to be blamed. On this picture, a moral ideal is Galilean – I'll go through the central features in turn.

First note that constructing a morally ideal agent is, on this picture, a practical project – determined by what best helps promote the cooperative goals. One needs to look both at what kinds of behaviour would be mutually beneficial, and in which cases criticizing deviance would be effective. For example, if a person were able to fly, that could prove incredibly useful – they could transport goods and rescue falling babies. However, criticizing someone for their inability to do so would achieve nothing. On the other hand, criticizing someone for giving in to temptation could well lead to improved behaviour in the future – our dispositions to resist temptation can be influenced. This might lead to a picture of a morally ideal agent which shared an actual agent's skills and knowledge but lacked (among other things) weakness of will.

---

<sup>19</sup> Note that I am concerned with Gibbard's ideas about the role of ethical concepts rather than with his expressivist semantics. I think it's perfectly possible to adopt this account of the role of moral and epistemic ideals, while assigning truth conditional content to sentences containing moral and epistemic terms

This suggests an ideal like the following (letting S be an agent and S\* be a model of S):

Uncompromising moral ideal: S\* is a moral ideal for S if

1. S\* has the same physical abilities and empirical knowledge as S
2. S\* desires to act in accordance with relevant rules and is immune from temptation.

A second important point is that there is room for multiple legitimate ideals, depending on our purposes. When working with the uncompromising ideal, any time an agent gives in to temptation they will be open to blame. Though useful, there is good reason to think that this ideal will not capture everything we might want from a moral ideal. We face a great deal of temptation, and this is not simply an imperfection to be criticized, but a feature that must be managed. There are more and less effective ways to deal with this: one can strive to avoid high stakes high temptation situations when one's will power is lower; one can form habits to reduce the amount of will-power required to do various arduous but necessary tasks etc.<sup>20</sup> It seems, therefore, like there should be an ideal that guides us in how do this effectively. Here's a suggestion (with S and S\* as before):

*Moral resource management ideal:* S\* is a moral ideal for S if:

1. S\* has the same capacities and knowledge as S and also suffers from temptation like S does.
2. S\* manages her will power effectively so as to best promote the cooperative goals – S\* will not allow herself to be vulnerable to temptation in a high-stakes situation; forms habits that allow her to minimize the force of temptation, etc.

Which ideal is more effective to hold people to will depend on the situation. For example, a schoolteacher who has limited time and is responsible for the actions of a large group of

---

<sup>20</sup> See, e.g., Holton (2009).

children may be better served by the uncompromising ideal. It is impossible for her to keep track of the level of temptation being faced by all 30 children and tailor her moral judgments accordingly. On the other hand, blaming children whenever they violate various general rules for behaviour, as dictated by the uncompromising ideal, would be manageable. Moreover, such a system would give general motivation for the children to behave appropriately, even if occasional violations were unavoidable – publicly criticizing one child will motivate the whole class to behave better. In contrast, a parent might be better served by the resource management ideal, since they would have greater knowledge of their child's situation and would be able to dedicate more time and energy to them individually – and moreover, coping with temptation effectively is a lesson *someone* needs to teach a child.

Finally, note that here too there is room for artefacts of the system – in this example, the agent meeting the uncompromising ideal would never exert will power (since she is immune from temptation) but that doesn't mean we actually aspire to do away with will power. This shows that moral ideals, on the ethical picture I've sketched, possess the three key features associated with Galilean ideals. Thus, we can make sense of a Galilean ideal playing a role within a normative practice – provided the relevant kind of normativity is understood in a suitably deflationary manner.

An important point to note is that I am not claiming that this is a plausible account of moral normativity. It is much too deflationary for most people's tastes – perhaps even for Gibbard himself. My aim was simply to show that it's possible to use a Galilean ideal to explain a normative phenomenon. My plan now is to show how we can give an analogous account of epistemic ideals and argue that in this case the account is plausible.

## **5. Epistemic Normativity**

As with the ethical case, I want to ground the normativity of epistemic concepts in mutually beneficial goals. I argued in chapter 2 that we have an interest in the doxastic states of our peers: in a cooperative society, it benefits everyone if everyone has, as far as possible, reliably true belief (or reliably accurate credences). This allows people to provide us with trustworthy testimony and to perform their duties effectively – I'll call

such beneficial features our epistemic goals. It would make sense, therefore, to have a practice of *criticism* of doxastic states that promoted our epistemic goals.

Given the discussion of previous chapters, a natural thought is that our epistemic concepts derive their normative force from this practice of criticism. Moreover, the idea has been suggested by many philosophers working on epistemic normativity, such as Brandom (2009), Craig (1990), Dogramaci (2012), and Sosa (1991). I will argue that this assumption leads to an attractive picture of epistemic ideals on which they are Galilean. My aim in this section is to show how this view of epistemic ideals follows from the assumption. I'll leave objections to the claim that epistemic normativity is grounded in these practical constraints until the next section.

Our working assumption amounts to the claim that epistemic concepts such as *rational* and *irrational* are used to express praise and blame, in order to further the epistemic goals – i.e. widespread accurate belief. Like with the moral case, epistemic ideals fit naturally within this practice. We look for a system of epistemic states that effectively promotes the epistemic goals – so that criticizing agents for deviating from the ideal and praising them for according with it will improve their epistemic practices.

As with the moral case, there is a practical aspect to selection of an ideal. We do not just want an ideal that has features that would promote the epistemic goals if realized, but one that it is useful to hold us to. This is why making the idealized reasoner empirically omniscient is a mistake, in just the way that making the ideal moral agent able to fly was. Criticizing us for not holding a strong belief on an empirical matter we had no evidence about would do no good – it's not something we could possibly get better at. Bayesianism, on the other hand, is an ideal based upon doing the best we can with what we've got to work with: the Bayesian ideal is to form the best epistemic states given your evidence.

It's much more plausible to think that criticizing people when their beliefs are not properly proportioned to their evidence will lead to improvement. This suggests that Bayesianism provides a significant ideal, given our working assumption:

Bayesian ideal:  $S^*$  is an epistemic ideal for  $S$  if:

1.  $S^*$  has the same evidence as  $S$

2. S\* possesses the set of credences best supported by her evidence.

However, like with the moral case, there seems to be a need for another ideal. There are many features of our psychology Bayesianism tells us nothing about, a number of which have been discussed at length in previous chapters. Here I'll focus on the following:

1. *Use of heuristics*: We don't always go through all the information we possess carefully in order to make judgments: sometimes we use short cuts and 'satisfice'.<sup>21</sup>
2. *Reasoning in working memory*: We can't update all our beliefs at the same time: we have to recall a few at a time and draw inferences in our working memory.<sup>22</sup>
3. *Compartmentalization*: Quite plausibly, our body of information is not an integrated whole but divided up into compartments (or possibly *modules*), which cannot be made consistent with each other in any straightforward manner.<sup>23</sup>
4. *Binary and partial belief*: The object of study for Bayesianism is a body of credences – or 'partial beliefs' – but there is psychological evidence that we possess full beliefs alongside these, as *sui generis* psychological states with a distinctive functional role.<sup>24</sup>

An uncompromising Bayesian might claim that these features of our cognitive structure are simply imperfections and thus the object of study for psychology rather than epistemology. However, this would be mistaken: these features can be employed in better and worse ways relative to the goal of arriving at accurate beliefs – in the same way as we saw for management of temptation in the ethical case. We should use heuristics when they work, but use explicit conscious assessment when the stakes are high and heuristics are likely to lead to error. We should make sure modules are coordinated on crucial

---

<sup>21</sup> See Kahneman (2011); Payne and Bettman (2004).

<sup>22</sup> See Cherniak (1990).

<sup>23</sup> This was the focus of chapters 3 and 4.

<sup>24</sup> See Holton (2013); Weisberg (ms).

issues. We should use binary beliefs when partial beliefs would lead to poor revisions of judgment. It is squarely within the realm of normative epistemology to systematize such rules. This suggests the following alternative ideal:

*Idealized epistemic resource management*: S\* is an epistemic ideal for S if:

1. S\* shares S's memory capacities, heuristic machinery, modular structure etc.,  
– along with her evidence.
2. S\* follows rules in the deployment of such features which tend to maximize reliability.

This ideal is clearly highly schematic – a lot of work must be done to spell out how these features of our psychological structure are best deployed, relative to our mutual epistemic goals. My aim is to make the case that investigating this ideal is a legitimate and pressing issue in epistemology. Important work has been done already (see, e.g., Cherniak (1990)) but I think it has been under-discussed. One reason for that, perhaps, is that it has been seen as a rival to Bayesianism – this is certainly how Cherniak presented it. However, once epistemic ideals are understood as Galilean, as I recommend, it can be seen that they are complementary projects, to be pursued in parallel. Thus investigating ideal epistemic resource management does not require us to abandon the fruitful applications of Bayesianism.

The ultimate goal here would be to form an analogue of the *decision theory* that Bayesian ideal is part of for the resource management ideal – so that we could evaluate *actions* as well as beliefs. Doing so would be of significant philosophical interest – however, this project would require investigation that goes far beyond what has been discussed. For present purposes, it is not necessary to fully elaborate the resource management ideal: in what follows, I only rely on the fact that such an ideal *exists* – this allows us to progress on our three puzzles.

A final point is, that, as with the moral case, there is room for artefacts of the system. In particular, on the Bayesian ideal there is no deductive reasoning, no concept acquisition, no binary but defeasible belief, the beliefs are completely integrated – these are not things we should aspire to. Thus, epistemic ideals, on this picture, meet all three

criteria for counting as Galilean.

## 6. Deeper Norms

I have shown how assuming that epistemic normativity is practically grounded leads to a picture on which epistemic ideals are Galilean – a view I hope is somewhat attractive. However, one might want to resist the initial assumption so that the project does not even get going. One might think that the practical project of promoting the mutual epistemic goals is all well and good, but claim that there is *more* to epistemic normativity than this. This is similar to the common objection to expressivist ethical views that they do not fully capture the force of our moral obligations.

I think, however, that this position is much less compelling in the epistemic case. For this argument to succeed, it must be that there is a single fundamental ideal that we have a robust normative connection to (i.e. there exists an Aristotelian epistemic ideal). Many defenders of Bayesianism take it to provide an ideal in this robust sense – I think this line of thought is clear in the passage from Smithies (2015) that I quoted in section 4. My aim in this section is to show that Bayesian ideal cannot be Aristotelian.

There are a number of ways to argue for an Aristotelian ideal. The most straightforward approach is to appeal to intuition. One might claim that it's intuitively clear that rationality is not a pragmatically grounded notion – and is answerable only to a single ideal. A more subtle argument appeals to the link between first-person deliberation and rationality. Consider the phenomenology of weighing the evidence when deliberating; it seems that one feels *pulled* by various epistemic reasons, and that this process shapes what judgments one makes. One might argue that there is a right way to respond to such 'pulls', and therefore introspection favours a robust concept of epistemic normativity.

I'll address these considerations in turn. First, I don't think appeal to intuition is particularly convincing. The intuitions are not as strong here as in the moral case and are disputed by a range of philosophers, who claim that epistemic normativity must ultimately rest on practical considerations.<sup>25</sup> This is compounded by the fact that most

---

<sup>25</sup> See, for example, Brandom (2009); Craig (1990); Dogramaci (2012); Sosa (1991).

people think that what we believe is not a voluntary matter – our beliefs are necessitated by our appraisal of the evidence rather than a matter of choice. Most philosophers who believe in a deep notion of *ethical* normativity want to tie it to free choice – we are only responsible for the things within our control – but without highly controversial assumptions about doxastic voluntarism, this consideration does not carry over to the epistemic case.

A second problem is that even if one grants such intuitive support for robust epistemic normativity, it's not clear that the intuitions support the Bayesian ideal over its rivals. Though we count inconsistencies in explicit reasoning as irrational, we do not intuitively take our compartmentalized psychological structure to be irrational. We don't, for example, judge athletes to be irrational in virtue of their being unable to articulate all of their practical knowledge. Thus, our intuitions seem to favour epistemic resource management as the robust ideal.

Now I'll turn to considerations concerning first-person deliberation. I am sceptical of such phenomenologically grounded intuitions – especially in such an obscure case where we're not dealing with the nature of, say, a visual representation, but with the cognitive 'phenomenology of the perceived force of an epistemic reason'. Again, though, the crucial point is that even if there is a notion of epistemic normativity linked to the first-person perspective, it cannot be the Bayesian notion. First-person deliberation is intimately connected with the non-Bayesian elements of our psychology discussed above. We employ binary beliefs, heuristics etc. when deliberating – and doing away with them is not an option. If, as I have been assuming, we are looking for a notion of an epistemic ideal that vindicates Bayesianism, it cannot be too closely connected to the first-personal deliberative process.

This points to a general issue that I think is worth emphasizing. Namely: one must be consistent the degree of robustness one ascribes to epistemic normativity. We saw in the discussion of *the happy medium puzzle* that the Bayesian enthusiast must take rationality to be an ideal in a relatively thin sense to allow for the theoretically fruitful machinery to get going, and to guard against the 'ought-implies-can' objection. If one does that, though, one can't appeal to intuitions that presuppose a robust notion of epistemic normativity when objecting to a Galilean conception of epistemic ideals. Because of this,

there is no room for a Aristotelian account of the Bayesian ideal.

## **7. Back to the Puzzles**

The Galilean conception of epistemic ideals provides a moderate defence of the Bayesian ideal. It grants that the ideal is a legitimate part of epistemology but also entails that it cannot provide the whole story. Further, it points to some potential problems with current practice, which contribute to difficulties in our puzzles. There are two pitfalls in particular I want to discuss: the first pitfall is that certain aspects of the ideal that are seen as significant may in fact be artefacts of the system; the second pitfall is that because the Bayesian ideal is not universally applicable, theorists may be trying to apply the ideal in cases it is not suited for. I'll show how watching out for them allows us to make progress on the puzzles we started the chapter with.

### **7.1 The Happy Medium Puzzle**

Recall our earlier statement of the happy medium puzzle: why should we think Bayesianism provides the normatively significant midpoint between our psychological structure and omniscience? The solution follows pretty directly from the discussion in section 5. There is *practical* reason for the Bayesian ideal to be a normatively significant midpoint once we see that it is a Galilean ideal – holding our peers to this ideal is an effective way to further our cooperative epistemic goals. Further, defusing the mystery, we can note that there is more than one significant midpoint – in some situations it is better to hold people to the resource management ideal.

### **7.2 Artefacts of the System and the A Priori**

It's well known that there are a number of artefacts of the system for orthodox Bayesian models. For example: there is no deductive reasoning – beliefs are updated only by conditionalization; it is not possible to revise one's attitude towards a proposition one initially has credence 1 or 0 in; and it is not possible to acquire new concepts. These are

not features of the ideal that have a normative force over us, and we're not in any danger of thinking otherwise.<sup>26</sup> However, there may be less obvious cases of artefacts of the system – and thus, we must be careful when drawing inferences about creatures like us from what we find in the ideal.

I think that Chalmers' defence of the a priori is an example of this. Recall that Chalmers aims to respond to the Quinian critique of the a priori, and leans heavily upon the Bayesian ideal to do so. His strategy, roughly speaking, is to identify a body of a priori propositions within the Bayesian model, and infer that these propositions are, in some sense, a priori for us. Crucially he requires these propositions to do significant theoretical work in his epistemology, metaphysics and philosophy of language – they are the building blocks he uses to 'construct the world'.<sup>27</sup> Above, I suggested it was puzzling to think that adopting the formal tools of Bayesianism could entail such significant epistemic commitments. I'll now argue that if epistemic ideals are Galilean, Chalmers' key inference is suspect – allowing us to defuse the puzzle.

I want to grant Chalmers' contention that there is a body of propositions that, within the Bayesian model, can reasonably be regarded as a priori. What I want to challenge is the idea that they apply to us in a significant way that makes them fit for the work Chalmers wants them to do. If the Bayesian ideal is Galilean, there is a possibility that the a priori propositions are artefacts of the system. Because of this, the complexity objection the traditional front-loading argument faced recurs.

The familiar examples from Quine, Putnam, et al. show that there will be no *simple* statements that meet Chalmers' a priori condition – all simple statements can potentially be undermined by empirical evidence. Indeed, if Williamson's claim about the indefinite extension of counter-examples is correct, the only thing that will qualify as a priori will

---

<sup>26</sup> Those who hold that there is an Aristotelian epistemic ideal are committed to thinking that the Bayesian ideal doesn't capture it completely – there must be a modified version that removes the artifact. Indeed work has been done to this end, looking at, e.g., additional rules of updating to add to the Bayesian ideal to deal with propositions one has credence 0 in.

<sup>27</sup> The introduction to Chalmers (2006) provides an eloquent summary of his ambitions: 'Carnap's link between meaning and modality, in conjunction with Kant's link between modality and reason, could be seen as building a Fregean link between meaning and reason. The result was a golden triangle of constitutive connections between meaning, reason, and modality... Two-dimensional semantics [Chalmers' preferred view] promises to restore the golden triangle.' It should be noted that the existence of a substantial body of a priori propositions, is essential to restoring 'the golden triangle.' His project of using such propositions to construct the world is carried out comprehensively in Chalmers (2012)

be the conditionals with maximal evidence statements for an antecedent. It doesn't seem plausible that these are things we have any attitudes towards whatsoever, never mind high conditional credences in across the board. I think, moreover, that this is not a defect the ideal corrects, but a deep fact about our epistemic structure. It's built into the Bayesian model that for any proposition  $p$  and evidence  $E$ , the agent has a well-defined conditional credence  $Cr(p|E)$ . Essentially, this is to assume that the rational agent works out in advance how to respond to every possible piece of evidence. For us, this is not at all the case: we don't even know what possible evidence we might receive, never mind know how to respond to it.

It is a central part of our epistemic practice that we face unforeseen evidence and work out what to do in response. Moreover, this is a rational process – one we can make a better or worse job of. We can reappropriate many of the traditional examples against the *a priori* to illustrate this. I think some of the most convincing ones are found in the work of Mark Wilson, as was discussed in chapter 1. Recall his (2014) discussion of our concept of *temperature*: the well-known claim that  $\text{temperature} = \text{mean kinetic energy}$  only works with gases. Liquids and solids have a frozen order which makes it impossible to get a sensible measurement of their mean kinetic energy. Complex modifications need to be made to talk about temperature in these case.

Wilson makes the plausible claim that this is not the kind of development that could be anticipated by our pre-theoretic conditional credences. Further, this doesn't look at all like the kind of case where we change the meaning of our term 'temperature' – it seems like a case in which scientists made some discoveries about temperature itself. Wilson (2006) presents a huge range of cases of this form, which together make a highly compelling case against the *a priori*.

Chalmers (2014), in his response to Wilson, argues that on the Bayesian model examples like the above discussion of temperature must either constitute updating by conditionalization or a change in meaning. He's right that this is true *of the Bayesian model*. However, if the Bayesian ideal is Galilean and if these examples do not look like they involve updating or meaning change in creatures like us, this gives us evidence that the features of the Bayesian model that entail this are artefacts of the system. Thus, we have good reason to think that the propositions Chalmers identifies as *a priori* are not

significant to us.<sup>28</sup>

### 7.3 Multiple Ideals and Higher-Order Evidence

The fact that there are multiple epistemic ideals means that no single ideal need give us the whole story about rationality. Therefore, it is not a defect of the Bayesian ideal if there are certain issues about rational thought that it is silent on (if these are not the cases the ideal was meant to deal with in the first place). If one is not mindful of this, one might feel compelled to distort the Bayesian model to attempt to make sense of examples that are best explained by the resource management ideal. In this section, I think that this issue may arise when it comes to the puzzle of higher order evidence.

Before getting into the specifics, I should note that I don't want to suggest that an ideal should *never* be modified to better deal with puzzling examples that it had not initially been intended to apply to – only that this is *sometimes* a bad idea. I'll illustrate this point by looking at frictionless ideals. As was mentioned above, such an ideal is well suited to offer predictions about an initial set of collisions, but is unable to predict the final resting places of objects set in motion. However, it is possible to predict the resting places by complicating the model to represent *interfering factors*, such as air resistance and friction. In this case, complicating the model is a good idea. One obtains an ideal that maintains much of the original's simplicity yet is able to make a new range of important predictions. If, on the other hand, we want to make predictions about the activities of the fundamental constituents of the objects we are examining, trying to complicate the frictionless ideal would be a terrible idea – instead we must abandon the Newtonian framework entirely and switch to a model from quantum mechanics.

I think that the same holds for the Bayesian ideal, and we will have to decide on a case by case basis whether complicating the ideal is an effective strategy for dealing with the relevant problem cases. I will suggest that when dealing with some of the problem

---

<sup>28</sup> One might wonder if some of the less radical elements of Chalmers' project survive. In particular, whether he is still able to give an account of meaning change in terms of changes in conditional credence. Nothing I've said here counts against this claim. However, I think that the examples we've been discussing show that we possess a much more limited range of conditional credences than are found in the Bayesian ideal. If these were the sole building blocks of meaning, I think it would be far more indeterminate than most philosophers would be happy with.

cases that seem to undermine level-bridging principles, we should make a clean break and switch to the resource management ideal.

Recall from section 2.3 that the puzzle of higher order evidence concerns the conflict between theoretical and intuitive considerations when evaluating level-bridging principles – these principles entail that the Bayesian model builds in perfect knowledge of one’s own evidence. There is strong theoretical motivation in favour of this principle, since it’s required in order to have *Conditionalization* as an update principle. On the other hand many cases seem best described as ones in which one’s higher-order evidence does not match one’s first order evidence. To handle this in the Bayesian frame work, one must either move to a more complicated framework, using something like conditionalization\*, or insist that the level-bridging principles still hold in these cases despite appearances (which entails that creatures like us are simply incapable of being rational in these situations).

An alternative possibility is that cases in which we have reason to doubt our own reasoning/evidence are better modelled by the resource management ideal. It seems to fit in well with the general system of strategies for deploying various cognitive tools in the situations in which they work best. If this were so, then it would seem plausible to think that the Bayesian ideal should be kept simple and elegant by maintaining strong level-bridging principles. We should stick with an easy to use formal tool and accept that it will not apply in cases where a subject has reason to be uncertain about their own evidence.

Perhaps, my scepticism is unjustified, and it will turn out to be possible to provide a modified Bayesian ideal that retains an acceptable degree of simplicity while providing useful results about cases of higher-order evidence. However, this is far from guaranteed. Moreover, the point I made in the previous section applies here too. Namely, that we should be careful when assessing the force of intuitions as applied to the Bayesian model. It might well be that we are in principle unable to gain complete access to our evidence, and it seems intuitively unacceptable to judge us irrational for this. However, this intuition seems to stem from a thick notion of epistemic normativity – one that does not apply to the Bayesian ideal. If strong level-bridging principles are essential to making the Bayesian ideal practically applicable, then we must keep them, no matter our intuitions.

## 8. (Final) Conclusion

I have been discussing the status of Bayesian ideals in epistemology. Regardless of what you make of my conclusions, I hope to have shown that the issue is more nuanced than is sometimes assumed: our choices with Bayesianism extend beyond ‘take it or leave it’. In particular one must decide what kind of ideal Bayesianism provides. At the least, I think I've made the case that viewing the Bayesian ideal as Galilean is a coherent position, so if one wants to use it in some more robust way in one's theorizing, an argument that it is Aristotelian must be provided.

I take one of the most important upshots of this chapter to be the connection between the robustness of epistemic ideals and the epistemic ‘ought implies can’ principles. I think one must be explicit in how closely one ties principles of rationality to our epistemic capacities – otherwise one runs into the kinds of problems I've discussed.

This chapter leaves a number of issues awaiting further investigation – most importantly, what kind of decision theory can be created with the resource management ideal. I think, though, that this is best left for another essay. Making progress on this requires a renewed look at the empirical literature – in particular relating it to the psychology of *desire*. The same is true of open questions from other chapters, such as what task-indexed belief can tell us about modes of presentation, and in particular the case of Mary the colour-scientist – here we need to know more about the ways in which our beliefs cluster into *compartments*.

What I think this tells us is that we should not attempt to go much further with pragmatism without pausing to see how our new conclusions relate to the concerns of naturalism and conceptual analysis. I said at the beginning of this thesis that pragmatism could not provide all the answers. My discussion of the practical role of attitude attribution placed new constraints on an account of the attitudes, and this might well spur progress using different techniques. It would be an unfortunate irony to persist with pragmatism when it is no longer useful.

## Appendix

# Knowledge and Gettier: The Plot Thickens

### 1. Introduction

Since Stanley and Williamson (2001) made their influential defense of intellectualism, Gettier cases have been thought to present a problem for the view. Intellectualists claim that know-how is a form of know-that: however, it's natural to think that know-that but not know-how is undermined in Gettier situations. That is, when a subject is in a Gettier situation, the mental states involved cannot be know-that but can be know-how. Thus, it is generally agreed that Gettier cases are a strike against intellectualism – though intellectualists think the blow is not decisive, and show how the intuitions can be explained away.

I will argue that Gettier cases in fact provide evidence for intellectualism when we fully consider the data. Standard discussions take Gettier situations and consider attributions of know-that and know-how in isolation. In these cases, it's natural to consider know-that attributions false and know-how attributions true. However, I will look at the case in which we attempt to attribute know-that and know-how at the same time. In this scenario it's not at all natural to offer the same judgments as before – we feel compelled to give both attributions the same truth-value. This is something the anti-intellectualist cannot explain. If know-how and know-that are different kinds of state, it should be perfectly acceptable to attribute the former while withholding the latter.

I argue that the best explanation of the data is provided by a contextualist intellectualism. On this view, 'knows' can pick out a range of relations depending on context. In some contexts it picks out a relation undermined by Gettier situations and in other contexts not. I will also look at the prospects for an invariantist intellectualism. On such a view, some of the intuitive attributions must be false, and one will have to give a

pragmatic explanation of our usage. I will suggest that such an account is considerably less promising than the contextualist alternative – though still in a better position than anti-intellectualism.

## 2. The Examples

I'll begin by presenting the relevant examples. The setup is based upon Cath (2011)'s case of 'the lucky light bulb'; however I've changed some of the details to make the intuitions as clean as possible.

Suppose Sophie is in a library dedicated to crackpot instructional guides. Sophie is unaware that this is a crackpot library and she approaches a shelf and pulls out a book. She happens to pull out a book that's not by a crackpot, it's the only one and it got put there by mistake. Sophie reads a sentence from the book, 'when changing a light-bulb one must first disconnect the power', and so forms the (true) belief that when changing a light-bulb one must first disconnect the power. If she'd picked up another book, she'd have read something false about changing a light bulb and believed that. So it seems plausible to say that Sophie does not know that when changing a light-bulb one must first disconnect the power. (This case is closely analogous to Goldman (1976)'s fake barn case, and Harman (1968)'s one true newspaper case.)

However, suppose that Sophie sits down and reads the whole book. It provides a complete and comprehensive explanation of how to change a light-bulb: first turning off the power, then pushing down gently on the old bulb and twisting it anti-clockwise to remove it etc. She takes all of this in and believes it. Moreover, she would be able to act on it, and so successfully change a light bulb if required. It therefore seems right to say that she knows how to change a light bulb.

So far, this is familiar and relatively uncontroversial, and is taken to present a problem for intellectualism – since Sophie appears to possess know-how but not know-that. However, things get more complicated when we try to fully articulate Sophie's predicament. In particular, the following seems contradictory (let's assume that first turning off the power is an essential stage in changing a light bulb):

Contradiction: When changing a light bulb, it's essential to first turn off the power; Sophie knows how to change a light bulb, but she doesn't know that one must first turn off the power.

Getting to grips with Contradiction is the primary task of this paper. Everyone should agree, I think, that it's a puzzling phenomenon. We are inclined to assert 'p' and to assert 'q', but are unwilling to assert 'p and q'. If we take these intuitive statements at face value then one of them must be false. But then we must explain why we make such statements. My primary task is to show how this complicates the project of coming up with an account of knowledge that fits with our intuitions. Though I argue for a particular solution (contextualism), this goal is secondary to the presentation of the puzzle. I'll now look at the possible responses.

### 3. Anti-Intellectualism

On a typical anti-intellectualist view, in the spirit of Ryle (1949), know-how and know-that are different kinds of states – the former is a certain kind of ability, and the latter a consciously accessible representation of a proposition.<sup>1</sup> According to this account, our inclinations to attribute know-how but not know-that when considering them in isolation are correct. However, when it comes to *Contradiction* we run into problems. The claim 'Sophie knows how to change a light bulb, but she doesn't know that one must first turn off the power' should be interpreted as attributing to Sophie (roughly) a practical ability to change light-bulbs, but lacking a consciously accessible representation of the proposition that one must first turn off the power. On this assumption, there should be no problem making such an attribution – it should be akin to saying 'she's strong but not tall'. However, this is not what it says at all.<sup>2</sup>

---

<sup>1</sup>My criticisms do not carry over to what George (2013) terms 'almost-reductionism' about know-how. Such a view, though technically not intellectualist, may well be able to offer a similar explanation to the one I endorse.

<sup>2</sup>Insofar as *Contradiction* can be interpreted as saying something non-contradictory, it does so by undermining the essentiality condition, and attributing to Sophie a risky light bulb changing method, done with the power on. One can avoid this complication by considering the claim 'Sophie knows how to safely change a light-bulb but she doesn't know that one must first turn off the power' which seems non-negotiably contradictory.

The anti-intellectualist must provide an explanation of why Contradiction, though true, is something we're unwilling to assert. One option would be to simply claim that we make a mistake – we're ignorant as to the truth-value of Contradiction. But if we are able to make the correct attributions in isolation, it's hard to see what would make us slip up when we bring them together. One could suggest that we implicitly believe (incorrectly) that know-how must be accompanied by know-that. We do not notice the inconsistency when making the attributions in isolation, but the implicit belief is 'activated' when we make the attributions side by side. This response implies that when we notice Contradiction we revise how we think about know-how and know-that, but this strikes me as highly implausible. Though we don't want to assert Contradiction, it doesn't illuminate a mistake in our theory of mind: we will happily go back to making the same kinds of know-how and know-that attributions in isolation as before as long as Contradiction is not salient.

The natural alternative to an error theory is to appeal to conversational pragmatics: that is, claim that though Contradiction is strictly speaking true, it conventionally implicates a contradiction by means of Gricean principles. I can't see how such an argument could work, however. If, as the anti-intellectualist claims, 'know-how' and 'know-that' pick out distinct properties, Contradiction is analogous to a sentence like 'Sophie is strong but not tall', which is perfectly assertible. Further, a canonical form of conversational implicature involves a sentence which, taken literally, expresses some completely useless information but allows the listener to infer that the speaker intends to communicate something useful (e.g. writing 'the student has excellent handwriting' in an academic letter of recommendation.) However, by the anti-intellectualist's lights, Contradiction expresses significant and correct information about Sophie, while asserting a contradiction is rarely useful in a conversation, so it's hard to see how a pragmatic explanation could lead the former to communicate the latter. Of course, the anti-intellectualist could argue that some aspect of the situation I have not mentioned yields the desired pragmatic effect, but I don't think we have good reason to believe such a thing existed unless it is presented to us.<sup>3</sup>

---

<sup>3</sup> An anonymous referee offers as an example of a sentence that can be true but seemingly imply a contradiction Moore sentences of the form 'I believe that p but not-p'. Though this provides proof of

#### 4. Intellectualism

Now I'll look at how intellectualists handle the situation. According to intellectualism, know-how attributions are equivalent to a certain kind of know-that attribution. The most well-known version of the view, defended by Stanley and Williamson (2001), says that sentence (1) has the underlying semantic form of sentence (2):

S knows how to  $\phi$

For some way of  $\phi$ 'ing  $w$ , S knows that  $w$  is a way for her to  $\phi$

Thus, an attribution of know-how just is an attribution of know-that. This fits nicely with our intuition that Contradiction is a contradiction, but uncomfortably with the attributions we make in isolation. Intellectualism seems to imply that either the intuition that Sophie possesses know-how or the intuition that she lacks know-that must be mistaken – since it's a consequence of the view that she must possess both or neither.

One might point out that the contents of the attributions considered are different: the content of the explicit know-that attribution is that one must first turn off the power, while the content of the know-how attribution is that  $w$  is a way of changing a light bulb. I don't think that this helps the intellectualist avoid the problem. By assumption, for any way  $w$  of changing a lightbulb, turning the power off is a stage of  $w$ . Therefore, if one knows how to change a lightbulb, the proposition one knows entails that one must first turn off the power. Moreover, this is not a case where appealing to a violation of closure seems promising. Sophie has explicitly read through all the stages of how to change a light bulb and believed all of them – there's no reason why she'd know the conjunction but not the conjuncts.

Another way the intellectualist could respond would be to appeal to modes of presentation. Stanley and Williamson (2001) suggest that know-how attributions

---

existence for such a form of implicature, we have no reason to think it also occurs with Contradiction. The phenomenon is connected with the relationship between the norms of assertion and the asserter's attitudes and hence first-personal attitude ascriptions. Nothing analogous seems apply for third-personal knowledge ascriptions, like the ones we are considering.

attribute knowledge under a so-called ‘practical mode of presentation’.<sup>4</sup> One might think this can help explain our intuitions. One could claim Sophie knows that one must first turn off the power relative to a practical mode of presentation but not relative to a theoretical mode of presentation – and thus our intuitive attributions are correct. The situation can be seen as analogous to our endorsement of ‘Lois Lane believes that Superman can y’ but not ‘Lois Lane believes that Clark Kent can y’.

However, this explanation is not satisfactory. First, it requires the controversial claim that knowledge under a practical mode of presentation is Gettier-compatible.<sup>5</sup> More importantly, this strategy only saves the attributions in isolation at the cost of undoing the explanation of *Contradiction*. If Sophie knows that one must first turn off the power under a practical mode of presentation, but not a theoretical one, then *Contradiction* should express this unproblematic truth. It should feel no more contradictory than the claim ‘Lois Lane believes that Superman can fly but she doesn’t believe that Clark Kent can fly’.

This shows how there’s not an easy way to accommodate this example for either position on intellectualism. In the next section I’ll show how a contextualist account can make sense of our intuitions.

#### 4.1 Contextualism

According to contextualism, ‘knows’ is a context sensitive term that picks out a relation between a subject and a proposition – with epistemic standards determined by the conversational context. Informally speaking, what it takes to ‘know’ something will vary, depending on how high the contextual standards are. Different versions of contextualism make different claims about what constitute standards for ‘knows’: it might be the relevant alternatives one must rule out, or the level of safety one’s belief must reach.<sup>6</sup>

---

<sup>4</sup> Opinion on the nature of practical modes of presentation varies. Stanley (2011) develops in detail the idea that they are analogous to first-personal modes of presentation, while Pavese (2014) claims they are analogous to a special kind of computer program. These differences will not be relevant to our purposes – neither will the issue of whether reference to a practical mode of presentation is a semantic or a pragmatic matter.

<sup>5</sup> Stanley (2011) explicitly rejects this claim, though I think Cath (2014) can be interpreted as defending it.

<sup>6</sup> See Rysiew (2007) for an outline of the various positions. Additionally, there are other non-invariantist theories of knowledge: relativism and expressivism. I will not discuss them here, but I think it should be

I will remain neutral on the nature of epistemic standards. All that's needed for present purposes is that in some contexts standards are high enough that 'knows' picks out a relation undermined by Gettier situations while in other contexts standards are weak enough that it picks out a Gettier-compatible relation. Although such a claim is not explicitly endorsed by most contextualists, it is compatible with their accounts – it is tentatively endorsed by Hawthorne (2000).<sup>7</sup> We then must show that in a context where we make explicit know-that attributions, the term typically picks out a Gettier-incompatible relation, while in contexts where we make know-how attributions, the term typically picks out a Gettier-compatible relation. This entails that a typical explicit know-that attribution to Sophie will be false, while a typical know-how attribution to Sophie will be true – thus it fits with our intuitions.

The really attractive feature of this view is the elegant treatment it provides of *Contradiction*. In general, if a context sensitive term is used multiple times in a single sentence, it will have the same reading throughout. One can't typically say, 'Kobe Bryant is tall and LeBron James is not tall' to communicate that Kobe is tall for a human while LeBron is not tall for a basketball player (despite LeBron being taller than Kobe). Therefore, though explicit know-that attributions used in isolation pick out a Gettier-incompatible relation, and know-how attributions used in isolation pick out a Gettier-compatible relation, when they are used together they must pick out the same relation. Thus, *Contradiction* really is a contradiction on this version of contextualism.<sup>8</sup>

The key challenge for the contextualist is explaining our intuitions when making at-attributions in isolation. We need to show how the context in which an explicit know-that attribution is made provides Gettier incompatible epistemic standards while the context in which an explicit know-how attribution is made provides Gettier compatible standards. I

---

clear that the contextualist strategy I employ will also be workable for these positions.

<sup>7</sup>One might object that such a move makes contextualism significantly less plausible, since there is something especially significant about the Gettier-incompatible relations, and thus 'knows' must only refer to them if it is to be theoretically significant. Perhaps this is so – though it's by no means uncontroversial. I will put this issue to one side, however, since I am only concerned with which account best deals with cases like Sophie, not with which account of knowledge is on balance most plausible.

<sup>8</sup>It's not an exceptionless law that context sensitive terms get the same reading throughout a sentence; it's just the default. Dorr (2014) explores ways in which one can get intra-sentence context shift with appropriate work – often by employing emphasis. As predicted by Dorr's findings, one can get a true reading of *Contradiction* if one places sufficient emphasis on the second 'know' – providing further evidence for the contextualist account.

will sketch an account of how context provides epistemic standards that allows for this. Whether such an account is ultimately correct depends on broader issues about context in philosophy of language, so my endorsement of this view must be tentative, pending a more complete account of conversational pragmatics.

It's plausible to think that knowledge attributions pick out cognitive achievements – epistemic features that are in some sense desirable.<sup>9</sup> The contextualist will say that the kind of achievement required for 'knowledge' is determined by the demands of the context – what standards for achievement are salient. Thus in everyday contexts, reliable perception will be a salient achievement, while in a philosophical conversation the achievement of ruling out exotic possible worlds may be salient.

Consider what cognitive achievement is salient when attributing know-how. A crucial point is that the thing known – that  $w$  is a way of  $\phi$ 'ing – is incredibly complex, so that representing it is not at all straightforward. Speaking loosely, one must not simply learn one thing but learn many, and grasp how they relate to each other – notice how Sophie had to read and retain the information written in a book chapter, not a single sentence, to learn how to change a light bulb. Thus, for know-how attributions, there is a salient cognitive achievement in representing the relevant information, regardless of the reliability etc. of the process that led to the representation. This predicts that the relation picked out by such attributions would not be undermined by Gettier phenomena.<sup>10</sup>

A typical explicit know-that attribution does have a single, relatively simple proposition as its content – e.g., that one must first turn the power off – such that representing it is not a significant cognitive achievement. Thus, the salient achievement is going to concern the justification. This predicts that the relation picked out by such attributions will be Gettier-incompatible.

This illustrates the pragmatic process the contextualist requires. We can further get a better grip on it by looking at the predictions it makes for other kinds of know-wh attributions. I'll argue that my preferred view is well placed to explain a range of cases in a systematic manner.

---

<sup>9</sup> See Sosa (1991).

<sup>10</sup> Note that it is the making of a knowledge attribution that alters the context in which the statement is to be assessed such a process is familiar from Lewis (1979).

## 4. 2 Know-wh

Suppose Sally is in a library dedicated to the work of crackpot historians. She is unaware that this is a crackpot library and she approaches the shelf and pulls out a book concerning the military conflicts of Native Americans. She happens to have pulled out a book that's not by a crackpot, it's the only one and it got put there by mistake. She reads a sentence from the book 'the Comanche used horses when waging war', and so forms the (true) belief that the Comanche used horses when waging war. If she'd picked up another book, she'd have read something false about the Comanche and believed that. So it seems plausible to say that Sally does not know that the Comanche used horses when waging war.

However, suppose Sally sits down and reads the whole book. It provides a comprehensive and correct explanation of why the Comanche dominated the southern plains: 'because they incorporated horses in their culture and adapted their warfare strategies accordingly. . . [they had] access to feral horses as well as theft from settlers...' Sally takes all this in and believes it. It now sounds right to say that Sally knows why the Comanche dominated the southern plains. And again, it sounds contradictory to say 'The key reason why the Comanche dominated the southern plains is that they used horses when waging war; Sally knows why the Comanche dominated the southern plains; but Sally does not know that the Comanche used horses when waging war.'<sup>11</sup>

We can see, therefore, that the same pattern of intuitions holds for know-why as for know-how. This is predicted by our account of how standards for knowledge are determined in context. Know-why requires representing a range of propositions that together make up an explanation and this is in itself a cognitive achievement – it makes salient epistemic standards that are Gettier-compatible.

However, consider a second case – a twist on a familiar example. Steve is driving through fake barn country where, unbeknownst to him, the countryside is littered with fake barns. He looks out the window at something that looks just like a barn, and by lucky coincidence it is – so Steve forms the true belief that there is a barn in front of him.

---

<sup>11</sup>This example is inspired by Kvanvig (2003)'s Comanche example – though he uses it to shed light on understanding.

Does Steve know whether there's a barn in front of him? It's natural here to say no – one is no more tempted to attribute know-whether than to attribute know-that.

This shows that the pattern of intuitions does not generalize to know-whether. Again, this is a welcome result. Know-whether, just like know-that, concerns representation of a single fact so the representation alone is not a significant cognitive achievement. Thus, there is not a salient Gettier-compatible reading for know-whether attributions.<sup>12</sup>

I will leave my discussion of contextualism here. Obviously, I have not presented a completely thorough contextualist account, but I think the sketch above, show how it's possible to get a workable version of the view that deals with the cases correctly.

### 4.3 Invariantism

I'll now look at whether the success of contextualism can be carried over to an invariantist intellectualist account. If one is an invariantist, one claims that every knowledge attribution of the form 'S knows that p at time t' expresses a single proposition with an absolute truth-value. A proposition such as the one attributing knowledge to Sophie is either Gettier-compatible or it isn't. Indeed, for traditional 'insensitive' invariantist accounts the knowledge relation is Gettier incompatible across the board.

However, more flexibility is allowed by Interest Sensitive Invariantism. On this view, the standards for knowledge may vary from subject to subject, depending on their practical concerns. However, this does not help in our case since the view is still entails that for a single knowledge attribution 'S knows that p at t' expresses a single proposition – so it must ascribe the know-that and know-how ascriptions we make to Sophie the same truth value.

Thus, for the invariantist, either the know-how attribution or know-that denial in the case of Sophie is mistaken. One must, therefore, explain why we make certain false

---

<sup>12</sup> A further question is whether the pattern generalizes to other types of know-wh: e.g. know-where and know-who. I won't explore them here since I think we've already said enough to motivate my contextualist account. I suspect, though, that they will be borderline cases, since they tend to be between know-why and know-whether in terms of the representational complexity involved – and thus whether the intuition pattern applies will depend on the details of the case.

attributions in such cases. It is implausible to attribute this to error on our part for the same reasons discussed with anti-intellectualism – even after seeing Contradiction we do not feel we were mistaken. Therefore, the invariantist must provide a pragmatic explanation.

What theory one offers will depend on whether one's invariantist account is Gettier-compatible or not. If it is compatible, one has to provide a pragmatic explanation of why we tend to deny explicit know-that attributions in Gettier situations. If it is not, one has to explain why we make know-how attributions in Gettier situations. It is hard to assess such a proposal in the abstract; we need to look at specific accounts of what the pragmatic process might be and see whether they are plausible.<sup>13</sup> The only worked-out approach I'm aware of is provided by Stanley, who takes know-that to be Gettier-incompatible. He argues as follows:

[W]hen we ask whether John knows how to ride a bicycle, we are typically only interested in whether, were John to set off on a bicycle guided by his belief about how to ride a bicycle, he would successfully be able to achieve his goal (perhaps of getting us chips). The reason we do not hear many ascriptions of knowledge-wh as Gettier susceptible is not because they are non-propositional. Rather, it is because the pragmatics of situations in which we ascribe knowledge-wh often places the focus on true belief, rather than justification. (Stanley, 2011, pp. 180-181)

Similarly, one could provide a pragmatic account that put into the pragmatics my ideas about the salient cognitive achievement being the focus (Stanley's account as it stands is in danger of over-generalizing to know-whether). One might think this suggestion is just as good as contextualism, given that it is putting in the pragmatics precisely what contextualism puts in the semantics.

However, this appearance is misleading since for a pragmatic account, the entire process has to be determined by general conversational principles. That is, such pragmatic principles imply that when one attributes a Gettier-incompatible state and a less demanding cognitive achievement is salient, one communicates that the subject

---

<sup>13</sup>cf. George (2013).

satisfies the less demanding state roughly, we communicate that the subject possesses mere true belief. This, however, is highly speculative, and Stanley doesn't provide sufficient evidence to motivate it. There are certainly many superficially analogous cases where the implicature does not occur. For example, in a context in which being over 5'10" is a salient achievement, saying 'Sophie is over six feet tall' does not generally communicate that Sophie is merely over 5'10". We are owed an explanation of why Gricean maxims trigger the implicature in the know-how case but not in these analogous examples.

Things get tougher when we consider Contradiction. For this statement to appear contradictory, the know-how attribution must no longer implicate mere true belief in this context otherwise it would make the quite reasonable claim that Sophie possesses a true belief but lacks appropriate justification. One might argue that once an explicit know-that attribution is made, justificatory standards are made salient, blocking the implicature of mere true belief. However, this again is worryingly speculative. When hearing Contradiction, one could equally well infer from the fact that know-how attributions typically implicate mere true belief, that the reasonable proposition was what the speaker was attempting to communicate. Given that pragmatic reasoning aims to lead to reasonable interpretations of utterances, this seems like a natural option.<sup>14</sup>

Perhaps, with further work, a Gricean explanation of the data will be found; however I am skeptical. As we've seen, our pattern of language use is complex and systematic. For now, therefore, we should take such examples as a mark in favour of contextualism.

---

<sup>14</sup> An analogous example occurs with 'or' – if we assume that the inclusive reading is semantic whilst the exclusive reading is sometimes pragmatically implied. You say 'In the battle John will die or Peter will die'; I reply 'it's not the case that John or Peter will die but that John and Peter will die'. We naturally interpret the 'or' in my statement as exclusive to avoid making it a contradiction. This is an analogous process to weakening the know-how attribution to make Contradiction consistent

## Bibliography

- Adams, M. P. (2009). Empirical evidence and the knowledge-that/knowledge-how distinction. *Synthese*, 170(1), 97-114.
- Alter, T., & Walter, S. (2006). *Phenomenal concepts and phenomenal knowledge: New essays on consciousness and physicalism*. Oxford University Press.
- Anderson, C. A. (2007). Belief perseverance. *Encyclopedia of Social Psychology*. Thousand Oaks, CA: Sage, 109-110.
- Arpaly, N. (2003). *Unprincipled virtue: An inquiry into moral agency*. Oxford University Press.
- Ball, D. (2009). There are no phenomenal concepts. *Mind*.
- Beck, J. (2012). The generality constraint and the structure of thought. *Mind*, 121(483), 563-600.
- Berker, S. (2013). Epistemic teleology and the separateness of propositions. *Philosophical Review*, 122(3):337-393.
- Berker, S. (2008). Luminosity Regained. *Philosopher's Imprint*, 8(2).
- Bianchi, D. (ms) Know How is Action-Indexed Information.
- Block, N. (1997). On a Confusion About a Function of Consciousness. In *The Nature of Consciousness: Philosophical Debates*, eds. N. Block, O. Flanagan and G. Guzeldere. Cambridge: MIT Press.
- BonJour, L. (1998). *In defense of pure reason: A rationalist account of a priori justification*. Cambridge University Press.
- Boyd R. (1991) Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds. *Philosophical studies* 61(1), 127-148
- Braddon-Mitchell, D. and Jackson F. (2007). The philosophy of mind and cognition.
- Brandom, R. (2009). *Articulating reasons: An introduction to inferentialism*. Harvard University Press.
- Bratman, M. (1987). Intention, plans, and practical reason. *Center for the Study of Language and Information*.
- Brownstein, M., & Madva, A. (2012). Ethical automaticity. *Philosophy of the Social Sciences*, 42(1), 68-98.

- Carr, D. (1979). The logic of knowing how and ability. *Mind*, 88(1), 394-409.
- Carruthers, P. (2006). *The architecture of the mind*. Oxford University Press.
- Cath, Y. (2011). Knowing how without knowing that. *Knowing How: Essays on Knowledge, Mind and Action*, 113-35.
- Cath, Y. (2014). Revisionary intellectualism and Gettier. *Philosophical Studies*, 172(1):7-27.
- Chalmers, D. J. (2006). The foundations of two-dimensional semantics. *Two-dimensional semantics*, 55-140.
- Chalmers, D. J. (2011). Revisability and conceptual change in. *The Journal of Philosophy*, 108(8):387-415.
- Chalmers, D. J. (2012). *Constructing the world*. Oxford University Press.
- Chalmers, D. J. (2014). Intensions and indeterminacy: Reply to Soames, Turner, and Wilson. *Philosophy and Phenomenological Research*, 89(1):249-269.
- Cherniak, C. (1990). *Minimal rationality*. Mit Press.
- Chrisman, M. (2007). From epistemic contextualism to epistemic expressivism. *Philosophical Studies*, 135(2), 225-254.
- Christensen, D. (2007). Does murphy's law apply in epistemology? self-doubt and rational ideals. *Oxford studies in epistemology*
- Christensen, D. P. (2004). *Putting logic in its place: Formal constraints on rational belief*. Oxford University Press
- Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, 67-90.
- Conee, E., & Feldman, R. (2001). Internalism defended. *American Philosophical Quarterly*. 38(1), 1-18
- Craig, E. (1990). *Knowledge and the state of nature: An essay in conceptual synthesis*. Clarendon Press
- Das, N. (ms) Knowledge and Safety from Rational Defeat.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes. *Journal of personality and social psychology*, 81(5), 800.
- Davidson, D. (1975). Thought and talk. *Mind and language*, 7-23.
- Davies, M. (1989). Tacit knowledge and subdoxastic states. In A. George (ed.),

- Reflections on Chomsky.*
- Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy*, 68(4), 87-106.
- Dennett, D. C. (1989). *The intentional stance*. MIT press.
- Devitt, M. (2011). Methodology and the nature of knowing how. *The Journal of Philosophy*, 108(4), 205-218.
- Dogramaci, S. (2012). Reverse Engineering Epistemic Evaluations. *Philosophy and Phenomenological Research*, 84(3), 513-530.
- Dorr, C. (2014). Transparency and the context-sensitivity of attitude reports. *Empty Representations: Reference and Non-Existence*.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. MIT Press.
- Egan, A. (2008). Seeing and believing: perception, belief formation and the divided mind. *Philosophical Studies*, 140(1), 47-63.
- Elga, A. (2013). The puzzle of the unmarked clock and the new rational reflection principle. *Philosophical Studies*, 164(1), 127-139.
- Elga, A. & Rayo, A. (ms). Fragmentation and Information Access.
- Enoch, D. (2006). Agency, shmagency: Why normativity won't come from what is constitutive of action. *The Philosophical Review*, 169-198.
- Evans, G. (1982). *The varieties of reference*. J. McDowell (Ed.). Oxford: Clarendon Press.
- Fantl, J. and McGrath, M. (2009). *Knowledge in an Uncertain World*. Oxford University Press.
- Feldman, R. (2000). The ethics of belief. *Philosophical and Phenomenological Research*, 667-695.
- Fodor, J. A. (1975). *The language of thought*. Harvard University Press.
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.
- Fodor, J. A. (1987). *Psychosemantics*. The MIT Press.
- Fodor, J. A. (1990). *A theory of content and other essays*. The MIT press.
- Fodor, J. A. (2001). *The mind doesn't work that way: The scope and limits of computational psychology*. MIT press.
- Fodor, J. A. (2008). *LOT 2: The Language of Thought Revisited*. OUP Oxford.
- Frankish, K. (2004). *Mind and supermind*. Cambridge University Press.

- Frege, G. (1984). On sense and reference. *Translations from the philosophical writings of Gottlob Frege*, 126-151.
- Frigg, R. and Hartmann, S. (2006). Models in science. *Stanford Encyclopedia of Philosophy/*
- Gendler, T. S. (2008a) Alief and belief. *Journal of Philosophy*, 105(10), 634.
- Gendler, T. S. (2008b). Alief in action (and reaction). *Mind & Language*, 23(5), 552-585
- Gendler, T. S. (2012) Intuition, Imagination, and Philosophical Methodology: Summary. *Analysis* 72 (4): 759-764
- George, B. (2013). Knowing-`wh', mention-some readings, and non-reducibility. *Thought: A Journal of Philosophy*, 2(2):166-177.
- Gertler, B. (2011). Self-knowledge and the transparency of belief. *Self-Knowledge*. Oxford University Press.
- Gettier, E. L. (1963). Is justified true belief knowledge? *Analysis*, 23(6), 121-123.
- Gibbard, A. (1992). *Wise choices, apt feelings: A theory of normative judgment*. Harvard University Press.
- Glick, E. (2011). Two methodologies for evaluating intellectualism. *Philosophy and Phenomenological Research*, 83(2), 398-434.
- Goldman, A. I. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy*, 771-791.
- Greaves, H. and Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115(459):607-632.
- Greco, D. (2014). Could KK Be OK? *The Journal of Philosophy*, 111(4), 169-197.
- Greco, D. (2014). Iteration and fragmentation. *Philosophy and Phenomenological Research*.
- Greco, J. (2003). Knowledge as credit for true belief. *Intellectual virtue: Perspectives from ethics and epistemology*, 111-134.
- Halberstadt, J. B., & Levine, G. M. (1999). Effects of Reasons Analysis on the Accuracy of Predicting Basketball Games. *Journal of Applied Social Psychology*, 29(3), 517-530.
- Hannon, M. (2013). The practical origins of epistemic contextualism. *Erkenntnis*, 78(4), 899-919.

- Harman, G. (1968). Knowledge, inference, and explanation. *American Philosophical Quarterly*, pages 164-173
- Haslanger, S. (2012). *Resisting reality: Social construction and social critique*. Oxford University Press.
- Hawthorne, J. (2000). Implicit belief and a priori knowledge. *The Southern journal of philosophy*, 38(S1):191-210.
- Hawthorne, J. (2002). Deeply contingent a priori knowledge. *Philosophy and Phenomenological Research*, pages 247-269.
- Hawthorne, J., & Stanley, J. (2008). Knowledge and action. *The Journal of Philosophy*, 105(10), 571-590.
- Hájek, A. (2003). Interpretations of probability. *The Stanford Encyclopedia of Philosophy*.
- Henderson, D. (2009). Motivated contextualism. *Philosophical Studies*, 142(1), 119-131.
- Holton, R. (2009). *Willing, wanting, waiting*. Oxford University Press.
- Holton, R. (2013). Intention as a model for belief.
- Jackson, F. (1982). Epiphenomenal qualia. *The Philosophical Quarterly*, 32(127), 127-136.
- Jackson, F., & Pettit, P. (1990). In defence of folk psychology. *Philosophical Studies*, 59(1), 31-54.
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan
- Kornblith, H. (2002). *Knowledge and its Place in Nature*.
- Kosslyn, S. M. (1996). *Image and brain: The resolution of the imagery debate*. MIT press.
- Kvanvig, J. L. (2003). *The value of knowledge and the pursuit of understanding*. Cambridge University Press.
- Laurence, S. and Margolis, E. (2002). Lewis's Strawman. *Philosophical Quarterly*, vol. 52.206, pp. 22-55. January 2002
- Laurence, S., & Margolis, E. (2003). Concepts and conceptual analysis. *Philosophy and Phenomenological Research*, 67(2), 253-282.
- Levy, N. (2015). Neither fish nor fowl: Implicit attitudes as patchy

- endorsements. *Noûs*, 49(4), 800-823.
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of philosophical logic*, 8(1):339-359.
- Lewis, D. (1982). Logic for equivocators. *Noûs*, 431-441.
- Lewis, D. (1984). Putnam's paradox. *Australasian Journal of Philosophy*, 62(3), 221-236.
- Lewis, D. (1988). What experience teaches. In William G. Lycan (ed.), *Mind and Cognition*. Blackwell 29--57
- Lewis, D. (1994). Reduction of mind. In Samuel Guttenplan (ed.), *Companion to the Philosophy of Mind*. Blackwell 412-431.
- Lewis, D. K. (1986). *On the plurality of worlds* (Vol. 322). Oxford: Blackwell.
- Loar, B. (1990). Phenomenal states. *Philosophical perspectives*, 4, 81-108.
- MacFarlane, J. (2005). The assessment sensitivity of knowledge attributions. *Oxford studies in epistemology*, 1, 197-233.
- Machery, E. (2009). *Doing without concepts*. Oxford University Press.
- Mandelbaum, E. (2012). Against alief. *Philosophical Studies*, 1-15.
- Mandelbaum, E. (2014). Thinking is believing. *Inquiry*, 57(1), 55-96.
- Marcus, R. B. (1990). Some revisionary proposals about belief and believing. *Philosophy and Phenomenological Research*, 133-153.
- Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: cognitive ability, cognitive style, and learning preference. *Journal of educational psychology*, 95(4), 833.
- Millikan, R. G. (1989). Biosemantics. *The journal of philosophy*, 281-297.
- Neta, R., & Rohrbaugh, G. (2004). Luminosity and the Safety of Knowledge. *Pacific Philosophical Quarterly*, 85(4), 396-406.
- Pavese, C. (2015). Practical senses. *Philosophers' Imprint* 15 (29).
- Payne, J. W. and Bettman, J. R. (2004). Walking with the scarecrow: The information-processing approach to decision research. *Blackwell handbook of judgment and decision making*, 110-132.
- Peacocke, C. (1992). *A study of concepts*. The MIT Press.
- Pettigrew, R. (2013). Epistemic utility and norms for credences. *Philosophy Compass*,

- 8(10):897-908.
- Pinker, S. (1997). *How the Mind Works*. NY: Norton
- Powers, L. H. (1978). Knowledge by deduction. *The Philosophical Review*, 87(3), 337-371.
- Prinz, J. (2006). Is the mind really modular. *Contemporary debates in cognitive science*, 22-36.
- Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological review*, 88(1), 16-45.
- Quine, W. V. O. (1970). *Philosophy of Logic*. Cambridge, MA: Harvard University Press.
- Railton, P. (2009). Practical competence and fluent agency. *Reasons for action*, 81-115.
- Ramsey, F. P., (1927). Symposium: Facts and propositions. *Proceedings of the Aristotelian Society*, supplementary volumes, 153-206.
- Ramsey, F. P., General Propositions and Causality. (1990). in his *Philosophical Papers*, ed. D. H. Mellor (Cambridge: Cambridge University Press, 1990) pp. 145–63.
- Recanati, F. (2004). *Literal meaning*. Cambridge University Press.
- Rumfitt, I. (2003). Savoir faire. *The Journal of Philosophy*, 158-166.
- Ryle, G. (1949). *The concept of mind*. Routledge.
- Rysiew, P. (2007). Epistemic contextualism. *Stanford Encyclopedia of Philosophy*.
- Santorio, P. (2016). Nonfactual Know-How and the Boundaries of Semantics. *Philosophical Review*, 125(1), 35-82.
- Schoenfield, M. (2015). Conditionalization does not (in general) maximize expected accuracy. Manuscript.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive psychology*, 22(1), 36-71.
- Schwitzgebel, E. (2010). Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly*, 91(4), 531-553.
- Shoemaker, S. (2009). Self-intimation and second order belief. *Erkenntnis*, 71(1), 35-51.

- Shook, N. J., & Fazio, R. H. (2008). Interracial Roommate Relationships An Experimental Field Test of the Contact Hypothesis. *Psychological Science*, 19(7), 717-723.
- Sider, T. (2013). *Writing the Book of the World*. OUP Oxford.
- Smithies, D. (2012). The Mental Lives of Zombies. *Philosophical Perspectives*, 26(1), 343-372.
- Smithies, D. (2014). The phenomenal basis of epistemic justification. In *New Waves in Philosophy of Mind*, 98-124. Palgrave Macmillan UK.
- Smithies, D. (2015). Ideal rationality and logical omniscience. *Synthese*, 1-25.
- Snowdon, P. (2004). Knowing How and Knowing That: A Distinction Reconsidered. *Proceedings of the Aristotelian Society* 104(1), 1-29. The Oxford University Press.
- Soames, S. (2002). *Beyond rigidity: The unfinished semantic agenda of naming and necessity*. Oxford University Press.
- Sosa, E. (1991). Intellectual virtue in perspective. *Knowledge in perspective: Selected essays in epistemology*, 270-293.
- Srinivasan, A. (2015). Are we luminous? *Philosophy and Phenomenological Research*, 90(2), 294-319.
- Stalnaker, R. (1984). *Inquiry*. The MIT Press
- Stalnaker, R. (1991). The problem of logical omniscience, I. *Synthese*, 89(3), 425-440.
- Stanley, J. (2000). Context and logical form. *Linguistics and philosophy*, 23(4), 391-434.
- Stanley, J. (2005). *Knowledge and practical interests*. Oxford University Press.
- Stanley, J. (2011). *Know how*. Oxford University Press.
- Stanley, J. & Williamson, T. (2001). Knowing how. *Journal of Philosophy*, 98, 411-44.
- Stich, S. P. (1978). Beliefs and subdoxastic states. *Philosophy of Science*, 499-518. In his (2011) *Collected Papers: Volume 1* (page references to the latter).
- Tye, M. (2000). *The imagery debate*. Mit Press.
- Wallace, D. F. (2006). Federer as religious experience. *New York Times*, 20.
- Wallis, C. (2008). Consciousness, context, and know-how. *Synthese*, 160(1), 123-153.
- Weatherson, B. (2003). What good are counterexamples? *Philosophical Studies*, 115(1), 1-31.
- Weatherson, B. (2005). Can we do without pragmatic encroachment? *Philosophical*

- Perspectives*, 19(1), 417-443.
- Weisberg, J. (ms). Belief in psyontology.
- White, R. (2006). Problems for dogmatism. *Philosophical Studies*, 131(3):525-557.
- Williams, B. (1973). Deciding to believe. *Problems of the Self*, 136(151), 401-419
- Williams, J. R. G. (2007). Eligibility and inscrutability. *The Philosophical Review*, 116(3), 361-399.
- Williams, J. R. G. (2013). Lewis on reference and eligibility. *A Companion to David Lewis*, 367-381.
- Williams, M. (1996). *Unnatural doubts: Epistemological realism and the basis of scepticism*. Princeton University Press.
- Williamson, T. (2000). *Knowledge and its Limits*.
- Williamson, T. (2006). Can cognition be factorised into internal and external components. *Contemporary Debates in Cognitive Science*. Oxford: Blackwell.
- Williamson, T. (2007). *The philosophy of philosophy*. John Wiley & Sons.
- Williamson, T. (2015). *Acting on Knowledge*.
- Wilson, M. (2006). *Wandering significance: an essay on conceptual behavior*. Oxford University Press.
- Wilson, M. (2014). David Chalmers versus the boll weevil. *Philosophy and Phenomenological Research*, 89(1):238-248.
- Wilson, T. D., Dunn, D. S., Bybee, J. A., Hyman, D. B., & Rotondo, J. A. (1984). Effects of analyzing reasons on attitude-behavior consistency. *Journal of Personality and Social Psychology*, 47(1), 5.
- Yablo, S. (2005). The Myth of the Seven. In Mark Eli Kalderon (ed.), *Fictionalism in Metaphysics*. Clarendon Press (2005)
- Zimmerman, A. (2007). The nature of belief. *Journal of Consciousness Studies*, 14(11), 61-82.

