

## University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]



# **University of Southampton**

Faculty of Humanities

Philosophy

## **Transparency and Doxastic Self-Knowledge**

by

**Eleanor Gwynne**

Thesis for the degree of PhD

December 2019



## Abstract

It is widely accepted that the knowledge we have of our own beliefs is distinct from other kinds of knowledge. However, in general, accounts of self-knowledge which capture this distinctive nature also struggle to explain precisely how our knowledgeable self-ascriptions of belief are warranted. That is, they fail to explain what makes our self-ascriptions *count* as knowledge. In this thesis, I seek an account that can do justice to the special status of doxastic self-knowledge, while also explaining the epistemology of our self-ascriptions.

The distinctiveness of self-knowledge, as Evans famously observed, can be partly seen to consist in the following fact: a subject can learn about her own beliefs by reflecting on the *subject matter* of those beliefs, as opposed to her own mental states. For instance, reflecting on the weather can put me in a position to know whether I believe it will snow. In this sense, beliefs are said to be 'transparent' to the world.

On one interpretation, the transparency thesis captures a way in which doxastic self-knowledge can be *grounded*, insofar as it suggests that world-directed judgments provide warrant for knowledgeable self-ascriptions of belief. I propose that by giving a satisfying epistemic account of 'transparent' self-knowledge, we can simultaneously satisfy the two key aims of my project – to explain what makes doxastic self-knowledge distinctive, as well as what makes it count as knowledge at all.

I begin by considering three promising accounts of doxastic self-knowledge, each of which purports to explain the source of our warrant for transparent self-ascriptions. The first two accounts, advanced by Richard Moran and Matthew Boyle respectively, understand our entitlement to transparent self-knowledge as inextricably connected to rational agency. On these views, it is because our beliefs are products of our capacity to 'make up our minds' that reflecting on their subject matter can itself ground self-ascriptions. The third account I consider represents a departure from this way of thinking. Alex Byrne's inferential transparency account gives no special role to rational agency, construing doxastic self-knowledge as the product of an inference from *p* to *I believe p*. The account relies on this inference meeting externalist conditions to explain why it is conducive to rational, knowledgeable self-ascriptions.

I will argue that, while none of the above accounts convincingly capture the epistemology of transparent self-knowledge, the views proposed by Moran and Boyle get something right in understanding our entitlement to self-knowledge as connected to our capacity for rational agency. In the second part of the thesis, I outline an alternative account that seeks to preserve the attractive elements of these views. Perhaps surprisingly, the view I defend also incorporates an inferential element.

In my positive account, I claim that conscious judgment should be understood as a form of mental action, the *phenomenal* character of which makes available to me the premise *I judge that p*. From this premise I can infer to the conclusion that *I believe that p*. This conclusion, I argue, constitutes a knowledgeable self-ascription of belief.



# Table of Contents

<b>Table of Contents .....</b>	<b>i</b>
<b>Research Thesis: Declaration of Authorship .....</b>	<b>iii</b>
<b>Acknowledgements.....</b>	<b>v</b>
<b>0. Introduction.....</b>	<b>1</b>
0.1 Doxastic Self-Knowledge .....	1
0.2 Traditional Accounts of Doxastic Self-Knowledge .....	3
0.3 Transparent Doxastic Self-Knowledge .....	5
0.4 Structure of Discussion.....	13
<b>Part 1: A Critical Analysis of Three Transparency Accounts .....</b>	<b>17</b>
<b>Chapter 1 Agentialism.....</b>	<b>17</b>
1.1 The Transparency Method .....	18
1.2 The Agentialist Epistemology .....	21
1.3 A Note on Deliberation .....	22
1.4 Objections to Agentialism .....	24
1.5 Conclusion .....	36
<b>Chapter 2 Reflectivism .....</b>	<b>38</b>
2.1 The Reflectivist Account .....	38
2.2 Objections .....	45
2.3 Conclusion .....	57
<b>Chapter 3 Inferentialism .....</b>	<b>59</b>
3.1 The Doxastic Schema and BEL.....	59
3.2 Privileged and Peculiar Access .....	61
3.3 A Note on the Importance of Self-Verification .....	65
3.4 Objections and Responses .....	66
3.5 Conclusion .....	76
<b>Part 2: A Phenomenological Approach to Transparency.....</b>	<b>79</b>
<b>Introduction to Part 2 .....</b>	<b>79</b>
<b>Chapter 4 The Metaphysics of Judgment .....</b>	<b>82</b>

## Table of Contents

4.1	Judgment as an Occurrent Event .....	82
4.2	Judgment as Belief Formation .....	84
4.3	Judgment as Mental Action .....	85
4.4	Objections and Responses to the Judgment as Action Thesis.....	93
4.5	Conclusion.....	104
<b>Chapter 5 Phenomenal Awareness of Judgments .....</b>		<b>107</b>
5.1	Consciousness and self-awareness.....	107
5.2	Judgment-Mineness.....	108
5.3	Judgment-mineness as constitutively grounding self-knowledge of judgments ..	112
5.4	The Epistemology of Transparent Doxastic Self-Knowledge .....	118
<b>Conclusion</b>		<b>133</b>
<b>References</b>		<b>137</b>



# Research Thesis: Declaration of Authorship

Print name: Eleanor Gwynne

Title of thesis: Transparency and Doxastic Self-Knowledge

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signature:

Date:



# Acknowledgements

I begin by thanking the AHRC for funding this project through the SWW Doctoral Training Partnership. I also thank the Royal Institute of Philosophy, who provided a bursary in support of my final year of research.

I thank Conor McHugh and James Stazicker, for their guidance, empathy, and patience during the writing of this thesis. Both have been extraordinarily generous supervisors. I thank Will McNeill for stepping in to take on supervisory duties at the end of the writing process. His enthusiasm and encouragement in these final stages were much appreciated. I also thank Giulia Felappi for her kindness and support as Doctoral Programmes Director.

I thank the wider Southampton philosophy department. Many members of staff have helped shape the contents of the thesis through their careful and considered feedback. Thanks are also due to Brie Gertler (University of Virginia), who graciously took the time to provide incredibly detailed written feedback on an early draft of chapter 1. Additionally, I am grateful to the administrative staff at the University of Southampton for their support over the course of my candidature.

I feel privileged to have been part of an inclusive PGR community during the writing of this thesis. The process would have been a far lonelier one without it. Special thanks go to Sophie Keeling, with whom I shared many fruitful discussions.

I would also like to thank my friends and family for their support, especially my parents, Jane and Val, and my Nan, Margaret Gwynne. Finally, I thank Charlie Boddicker, my closest friend and the funniest, most well-read person I know, for being a constant source of love and encouragement.



## 0. Introduction

### 0.1 Doxastic Self-Knowledge

A person is in a position to know all manner of things about herself. She might, for instance, know that she is afraid of open water, that she loves salty foods, or that her right leg is sore. She probably knows many of her own character traits, such as her tendency to procrastinate, and facts about her personal history, like the name of her first pet. Perhaps she knows that she intends to watch *Carnival of Souls* tonight, or that she desires a hot cup of coffee. She is also in a position to know her own beliefs. Some of these beliefs might be relatively trivial, like the belief that her coffee has gone cold, or that it is snowing outside. But she can also have knowledge of more important beliefs, such as whether she believes her career is fulfilling, or whether she believes that one should vote for a particular political party. All of these examples represent the kind of thing we might be talking about when we talk about self-knowledge.

If there is one question central to philosophical debate on the topic of self-knowledge, it is the question of whether it is different from other forms of knowledge, and if so, what this difference consists in. Of course, there are things one learns about oneself in much the same way as others do, bodily attributes being the most obvious example of this. If I want to learn my height I will have to get the tape measure out – I have no unique way of learning facts like this about myself.

However, when it comes to learning of my attitudes, such as my beliefs, desires, and intentions, it is less obvious that I will need to employ the same kinds of empirical methods required to learn about my bodily attributes. While my friend might rely on observing my facial expressions, body language, and tone of voice to provide information on what I am thinking and feeling, I myself need not rely on these sorts of indicators to tell me about my own psychology.

This thesis will be concerned with exploring the idea that we have distinctive self-knowledge of *beliefs* specifically. I restrict my discussion in this way for two interconnected reasons:

#### *The distinctiveness of the agent-belief relationship*

If any kind of self-knowledge is going to be distinctive, doxastic self-knowledge is a good candidate. This is partly due to the particular relation a subject bears towards her own beliefs. More obviously than other mental phenomena, beliefs are tied to the capacity for active reasoning. A human subject appears to be at least partially responsible for forming her beliefs, hence why we would ordinarily treat her as accountable for them. Her beliefs

are things she herself plays a role in bringing about and maintaining, as opposed to things that merely happen to her. Accordingly, we regard the relationship a subject has with her own beliefs as quite different from that which she has with the beliefs of others. The fact that a subject has this connection to her beliefs naturally invites the question of whether this puts her in a unique position when it comes to knowing them, in contrast to the position she is in with respect to knowing other people's beliefs.

### *The predominance and importance of doxastic self-knowledge*

The vast majority of literature on self-knowledge has tended to focus on doxastic self-knowledge, meaning this is where we find the most fully developed accounts. This, I think, can be explained partly by the above consideration, but more generally by the fact that belief represents a fundamental feature of our mental lives. What a subject believes influences and even dictates her other attitudes. Belief also plays an important explanatory role for action and behaviour. For a subject to know what she believes puts her in a position to understand and evaluate her own take on the world. It puts her in a position to authoritatively report this, and to revise it where appropriate. Knowing what one believes is thus paramount to being an effective agent.

Based on the above considerations, it seems we have good reason to suppose that self-knowledge of beliefs is both important and distinctive. I will assume as much in this thesis. I thus begin with the thought that satisfying account of doxastic self-knowledge should be able to capture this distinctive nature. However it should also be able to tell a convincing epistemological story about how such knowledge is warranted. In other words, we should expect an account of doxastic self-knowledge to explain the following:

- 1) Why a subject's knowledgeable beliefs about her own doxastic states are different from other knowledgeable beliefs (*distinctiveness*).
- 2) Why these second-order beliefs count as knowledge (*epistemic warrant*).

I will argue that existing views generally fail to give satisfactory explanations for either 1 or 2. Broadly, accounts that capture the distinctiveness of self-knowledge struggle to explain how it is warranted, and those that give a straightforward explanation of the epistemic warrant for doxastic self-knowledge fail to capture its distinctive nature. I contend that we can do justice to both 1 and 2 if we can establish a plausible epistemology for a particular form of doxastic self-knowledge.

In the next section of this chapter I will discuss two traditional accounts of doxastic self-knowledge: the inner-sense theory, and Rylean inferentialism. Both of these accounts, in one way

or another, reject that self-knowledge has a distinctive epistemic status. In the case of the inner-sense model this is because self-knowledge is construed as similar to perceptual knowledge. Inferentialism on the other hand rejects that doxastic self-knowledge is the product of a perception-like mechanism, and instead claims that it is the product of the same inferential procedures used to learn of others' beliefs. By modelling self-knowledge on other knowledge, both views provide clear accounts of the epistemic warrant for knowledgeable self-attributions of belief. But, I argue, in doing this they are also unable to capture its distinctiveness fully.

In section 3 I discuss the notion of 'transparency', a particular distinctive feature of doxastic self-knowledge that traditional views cannot explain. Self-knowledge of a belief can be thought of as transparent insofar as it is acquired by thinking about the subject matter of that belief. I note three promising accounts which do explain this feature, those advanced by Moran (2001, 2003), Boyle (2011), and Byrne (2011, 2018) respectively. These accounts, I contend, explain the distinctiveness of doxastic self-knowledge, thus fulfilling the first of my above desiderata, but struggle to convincingly explain the source of epistemic warrant for doxastic self-knowledge, and so do not fulfil the second.

Finally, in section 4 I briefly outline my own proposal for an alternative account. This account has similarities with that advanced by Boyle, but incorporates elements of what is known as the 'reasons' account (McHugh (2012), Peacocke (1998, 1999)), as well as O'Brien's 'agent's awareness' account of self-knowledge of actions (2007), and literature on phenomenal self-awareness, in particular the phenomenal feature of 'mineness' (Kriegel 2003, 2009; Zahavi 2000, 2005, 2011, and 2014; co-authored 2015).

## **0.2 Traditional Accounts of Doxastic Self-Knowledge**

In this section I outline two historically orthodox views on doxastic self-knowledge. These accounts are standardly well-positioned to explain the epistemology of doxastic self-knowledge, since they construe it as essentially similar to other forms of knowledge. However, as a result of aligning doxastic self-knowledge so closely with other knowledge, these accounts also struggle to capture its distinctiveness.

### **0.2.1 The Inner-Sense Model**

The inner-sense model is the standard historical conception of self-knowledge. According to views that fall into this category, acquiring doxastic self-knowledge involves a form of inward-directed observation. This process is regarded as similar to ordinary perception, though the comparison is largely metaphorical. While ordinary perception involves receiving information through sensory

## Introduction

organs, there is of course no equivalent introspecting organ. Nonetheless, inner-sense accounts hold that self-knowledge of beliefs is acquired by way of an internal scanning mechanism that is comparable to sensory mechanisms.

Views of this sort represent a development of Locke's account of introspective self-awareness:

... the perception of the operations of our own mind within us, as it is employed about the ideas it has got; which operations, when the soul comes to reflect on and consider, do furnish the understanding with another set of ideas, which could not be had from things without... all the different actings of our own minds; which we being conscious of and observing in ourselves, do from these receive into our understandings as distinct ideas, as we do from bodies affecting our senses. This source of ideas every man has wholly in himself; and though it be not sense, as having nothing to do with external objects, yet it is very like it, and might properly enough be called internal sense... I call this reflection, the ideas it affords being such only as the mind gets by reflecting on its own operations within itself. (Locke 1689/1975: II.1.iv.)

The key ideas from this passage can be reconstructed as follows: self-knowledge of mental states (including beliefs) comes from reflecting upon, or observing, one's own mental states. When one engages in reflection upon one's own first order mental states, these states cause new second order states to be formed. These second order states are representations of the first order states, and are knowledgeable when they accurately represent their subject matter.

The reflective process by which Locke takes one to acquire knowledge of one's own mental states is similar to Locke's view of sensory perception. On this view, mental representations of external objects are caused by those objects, and one can only have second-order representations of one's own mental states, and thus knowledge of these states, by observing them.

The ideas expressed by Locke notably re-appear in the inner-sense account advanced by Armstrong, who conceives of self-knowledge, including doxastic self-knowledge, as the product of a 'self-scanning process' (1993: 324). On this view, mental states cause representations of themselves, and one need not be aware of this causal connection in order to have immediate, i.e. non-inferential, self-knowledge on the basis of introspected representations.

### 0.2.2 Rylean Inferentialism

The inferentialist position advanced by Ryle is similar to the inner-sense model insofar as it also construes self-knowledge as similar to other knowledge. However, it rejects the thought that self-knowledge is acquired through a special introspective scanning mechanism.



According to Ryle (1949), doxastic self-knowledge is the product of the very same kinds of *inferential reasoning* employed to learn of other people's beliefs. It is produced by treating one's own behaviour and occurrences of inner-speech as *evidence* from which inferences about one's own beliefs can be drawn. Ryle's inferentialism thus denies the special epistemic status of doxastic self-knowledge by rejecting the thought that a subject can come to learn of her own beliefs in a distinctively first-personal way.

Ryle's view of doxastic self-knowledge, according to which self-knowledge is no different from other knowledge (apart from the fact that an individual is usually in a better position to gather evidence about herself than about other subjects), is among the less popular positions in the literature. It is now widely accepted that self-knowledge should be understood as having some kind of special status that sets it apart from other knowledge, though the details of this special status are widely debated. However, it should be noted that a minority of contemporary theorists defend views similar to Ryle's (Carruthers (2011), Cassam (2014)). These accounts are accordingly referred to as 'neo-Rylean' theories of self-knowledge.<sup>1</sup>

Neo-Ryleans agree that doxastic self-knowledge is essentially similar to knowledge of other people's beliefs. However, whereas Ryle restricts the evidence a subject can use to determine her own beliefs to facts about her behaviour and inner-speech, neo-Ryleans allow that other mental phenomena can constitute evidence for belief, including 'judgments, inner speech, dreams, passing thoughts and feelings' (Cassam, 2014: 138).

### 0.3 Transparent Doxastic Self-Knowledge

In this section I detail a particular distinctive feature of doxastic self-knowledge which the traditional views discussed in 0.2 are seemingly unable to account for – the feature of transparency.

#### 0.3.1 What is Transparency?

It is sometimes said that beliefs are 'transparent' to their subject matter, insofar as a subject can learn *what she believes about p* by reflecting on *p itself*. In other words, a subject can come to

---

<sup>1</sup> Byrne (2011, 2017) is also sometimes positioned as a neo-Rylean, for instance by Dorit Bar-On and Jordan Ochs (2018). This is understandable, given that he regards self-knowledge as inferential. However, the spirit of Byrne's account is markedly different from Ryle's. According to Byrne, doxastic self-knowledge *does* have a special epistemic status, and the model of inference through which we attain doxastic self-knowledge is explicitly *not* like ordinary forms of inference (in fact it is distinctively first-personal). For this reason, I here avoid presenting Byrne's account as a contemporary development of Ryle's.

## Introduction

have self-knowledge of her own belief by thinking about something other than the belief itself. Namely, the thing the belief is about.

To use the now well-worn example, it seems I can acquire knowledge of my belief about whether it is raining by reflecting on the weather outside, as opposed to reflecting on my own mental states. In answering the world-directed question of 'whether  $p$ ' I come to be in a position to answer the question of 'whether I believe that  $p$ '. To put it another way, the question of what I believe can be answered by engaging in the very same kind of thinking I would employ to answer the world directed question.

Outside of philosophical discourse, the significance of this observation may not be immediately obvious. Indeed, it may be seen as trivial that we should learn what we think about  $p$  by reflecting on  $p$ . However, if we are indeed in a position to learn about our own beliefs 'transparently', this suggests that self-knowledge is quite unlike other forms of knowledge. Ordinarily, the way to gain knowledge of something is by observing or reflecting upon that very thing, or by reflecting upon its effects. Given this fact, the notion that I can learn about my own doxastic states without the need to reflect upon these states or their effects starts to look quite striking.

This notion of transparency is most commonly associated with Gareth Evans, who is credited with giving the first explicit expression of the thesis. In an oft-quoted passage from *The Varieties of Reference*, Evans notes that when a subject goes about the task of self-ascribing a belief, she does so by attending to features of the world relevant to the subject-matter of that belief. In his famous example, answering the question of *whether I believe there will be a third world war* involves the very same kind of reflection employed to answer the question of *whether there will be a third world war*.

'[...]in making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me "Do you think there is going to be a third world war?", I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?"' (Evans, 1982: 225)

Importantly, the procedure described can only yield knowledge of a subject's own beliefs. It would not be conducive to gaining knowledge of other people's beliefs. I could not, for example, learn whether my friend believes there will be a third world war by reflecting on whether there will be a third world war. This captures one asymmetry between knowledge of one's own beliefs and knowledge of others' beliefs, what Byrne refers to as 'peculiar access' (2011, 2018).

The transparency thesis captures a fundamentally first-personal means of acquiring self-knowledge. To learn of one's own belief transparently need not involve the use of an internal scanning mechanism, nor the employment of inferential methods involved in learning of other peoples' beliefs.

### 0.3.2 Notable Transparency Accounts

Accounts of doxastic self-knowledge developing the idea outlined by Evans have been posed by Dretske (1994), Moran (2001), Fernández (2003, 2013), Boyle (2009, 2011), and Byrne (2005, 2011, 2018). I here consider the accounts I take to be most promising: those advanced by Moran, Boyle, and Byrne. The first half of this thesis (Chapters 1-3) consists of extended critical discussions of these three views.

#### 0.3.2.1 Moran's Agentialism

The accounts advanced by Moran and Boyle both fall into the category of views referred to as 'agentialist'<sup>2</sup>. However, for the purposes of distinguishing them I will reserve this term for Moran's account, and employ the more specific term 'reflectivist' for Boyle's view.

On the view Moran advances, a subject learns about her own beliefs 'transparently' by engaging in deliberation over what her reasons support.

...a first-person present-tense question about one's belief is answered by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world. (2001: 62)

So, to learn what she believes about  $p$ , a subject considers the reasons relevant to the question of whether  $p$ . In this sense, when the subject deliberates she treats the question of whether  $p$  as 'equivalent' to the question of whether she believes that  $p$ . If she decides that her reasons support that  $p$ , this puts her in a position to knowledgeably self-attribute the belief that  $p$ .

To be clear, the subject in this case can learn about her own beliefs by considering her reasons and *nothing else*. This is because, insofar as she is a *rational agent*, her reasons shape her beliefs. According to Moran's agentialism, any belief that is rational, in the sense that it conforms to the agent's reasons, is one that the agent could learn about transparently, i.e. by reflecting on the reasons relevant to the subject matter of the belief (2001: 107-8).

---

<sup>2</sup> Much of the terminology in this area, including 'agentialism', is due to Gertler (2017, 2011).

## Introduction

Importantly, the subject's world-directed judgment that *p* non-inferentially grounds her knowledge that *she believes that p*. That is, the subject need not make an inference from her judgment that *p* to the belief that *she believes that p* in order for her self-ascription of the belief that *p* to be warranted.

### 0.3.2.2 Reflectivism

Boyle's reflectivist account represents a development of the central ideas of Moran's agentialism. According to this view, the key to understanding why we can learn of our own beliefs transparently, i.e. by reflecting on their subject matter, lies in understanding the metaphysical nature of belief.

Reflectivism contends that believing that *p*, and believing that *one believes that p* constitute a single 'psychological condition' (2011:235), meaning whenever a subject has the belief that *p*, she necessarily knows that *she believes that p*. In a sense, self-knowledge is built into belief. The form of self-awareness constitutive of believing is referred to as 'tacit knowledge'. I take tacit here to mean *implicit*.

On my reading of the view, a subject knows that *she believes that p* purely in virtue of believing that *p* because believing is a form of mental agency, and exercises of agency are things of which a subject must be aware. More specifically, the belief that *p* is *constituted* by an on-going, active, commitment to *p* being true.

...we have privileged knowledge of [beliefs], not because we are in a specially good position to form second-order beliefs that reliably track their existence, but because their existence is normally constituted by our knowing assessment of the matter in question. (2011: 237)

We can restate the above thought in slightly different terms: to believe that *p* just is to be knowingly, actively engaged in affirming the propositional content that *p*. Explicitly self-attributing the belief that *p* on this view, or consciously manifesting one's self-knowledge, is a matter of *reflecting* upon one's active 'assessment' in the right kind of way, as it constitutes a belief one already has knowledge of.

So, the reflectivist position is committed to three core ideas: (1) tacit (implicit) knowledge that *one believes that p* is constitutive of believing that *p*, (2) 1 is true because believing that *p* is an exercise of mental agency constituted by a knowing, active assessment of *p* as true, (3) the implicit knowledge one has of one's own belief that *p* can be made explicit by reflecting upon this part of the composite state, or 'condition', described by 1.

*Prima facie*, reflectivism might not look much like a transparency account, given that it does not describe a transparency *procedure* as such.<sup>3</sup> However, it does defend a kind of metaphysical transparency: one can know what one believes about *p* by engaging in reflection upon *p* because one's knowing assessment of *p* constitutes the belief that *p*, and self-knowledge is constitutive of this belief.

Boyle's view can therefore be thought of as making a negative claim about transparent self-knowledge. Because believing that *p* just is to have transparent self-knowledge of that belief, there is no further procedure I must engage in to become aware that I believe that *p* when I have this belief.

### 0.3.2.3 Inferentialism

Byrne's inferentialist account offers a significantly different development of the transparency thesis from those we find in Moran and Boyle. This view does not conceive of transparent self-knowledge as being connected to rational agency in any important way. Nor does it construe this kind of self-knowledge as non-inferentially grounded.

According to the inferentialist transparency account, when a subject acquires self-knowledge transparently she makes an inference from a premise, *p* (some fact about the world), to the conclusion *I believe that p*. This model of inference is known as the 'doxastic schema' (Gallois, 1996). Inferentialism states that the doxastic schema is 'strongly self-verifying'; making the inference is meant to guarantee the truth of the conclusion, since one must believe a premise in order to infer from it. Moreover, it is argued that beliefs formed by reasoning in line with the schema are 'safe', in that they could not easily be false.

Inferentialism aims to explain two distinctive features of doxastic self-knowledge, characterised as 'privileged access' and 'peculiar access'. Privileged access refers to the fact that beliefs about one's own beliefs are more epistemically secure than beliefs about other people's doxastic states. Peculiar access, as previously noted, refers to the fact that there is a uniquely first-personal way of acquiring doxastic self-knowledge. Inferentialism is supposed to explain privileged access because the doxastic schema is strongly self-verifying and produces safe beliefs. It explains peculiar access because reasoning according to the doxastic schema would not be conducive to gaining knowledge of other people's beliefs.

---

<sup>3</sup> Indeed Boyle's view has clear similarities with Shoemaker's constitutivist position on doxastic self-knowledge, according to which believing that *p* constitutes knowing that *one believes that p*, provided the subject is sufficiently rational, and in possession of relevant concepts (Shoemaker, 1994).

### 0.3.3 The Problem of Epistemic Warrant

The transparency accounts outlined in 0.3.2 all face one crucial problem: they do not provide clear or plausible explanations for why transparently acquired beliefs about one's own beliefs are warranted such that they count as knowledge. Another way of expressing this is that they fail to explain one's *entitlement* to transparent self-knowledge. In the cases of Moran and Boyle, this is partly because they downplay the importance of epistemology in capturing the distinctive nature of doxastic self-knowledge. What is primarily important on these views is describing doxastic self-knowledge in such a way that accounts for the fact that beliefs are expressions of rational agency.

On one interpretation, Moran and Boyle are not attempting to give epistemic accounts of self-knowledge at all.<sup>4</sup> However, the agentalist and reflectivist views are attempting to advance real alternatives to more traditional epistemic conceptions of doxastic self-knowledge. As both Gertler and Roessler rightly point out, this involves advancing an explanation of how self-knowledge is warranted:

Agentialism [including reflectivism] takes self-knowledge to rest on epistemic entitlements or rights deriving from rational agency. It is thus a direct competitor to empiricism [i.e. traditional accounts], regarding the epistemology of self-knowledge. (Gertler, 2018: 94)

The stated aim of Moran's theory is to 'explain and vindicate avowal as a privileged form of knowledge of oneself' (2001, 134). More specifically, the idea of an entitlement to form second-order beliefs on the basis of one's 'explicit thinking' is absolutely central to the theory. (Roessler, 2013: 7)

In my critical analyses of Moran and Boyle's respective views, I will aim to show that neither succeed in giving an explanation of the epistemic warrant for doxastic self-knowledge that can rival the explanations of traditional views.

Byrne's inferentialism on the other hand is more obviously concerned with the epistemic features of self-knowledge, given its focus on explaining the special access subjects have to their own beliefs, and the idea that transparently acquired beliefs about one's own doxastic states are especially secure.

---

<sup>4</sup> Reed (2010) explicitly states this of Moran's account.

However, Byrne's view faces a different kind of issue, in that the particular inference it appeals to does not seem able to yield knowledge. As we will see, the inference captured by the doxastic schema must be either redundant, or else nonsensical (Boyle, 2011).

### 0.3.4 An Alternative Transparency Account

In now I outline my alternative account of doxastic self-knowledge, advanced in the second half of the thesis. The goal of my account is to show both what happens when one learns of one's own beliefs transparently, and why transparently acquired beliefs about what one believes are warranted so as to be knowledgeable. I will pay special attention to the issue of warrant, given that this is where most transparency accounts struggle.

Since my primary focus will be on understanding why 'transparently' acquired beliefs about one's own doxastic states are *epistemically warranted* such that they count as knowledge, I consider the following expression of the transparency thesis suitable for the particular goals I have in mind:

When a human subject judges that *p*, this judgment warrants the subject in believing that *she believes that p* in such a way that this belief would count as knowledge. Doxastic self-knowledge can thus be said to be 'transparent' insofar as it is grounded by a judgment about the *subject-matter* of a given belief.

My conception of the transparency thesis differs from Evans's original statement of the idea in that it shifts the focus away from the procedure employed to learn of one's beliefs transparently, and instead advances a more specific claim about epistemic entitlement to self-knowledge. To be clear, I do not regard my expression of the transparency thesis as definitive. Nor do I regard it as a refinement or improvement of other expressions. Rather, the version of the thesis I advance is intended to express the particular interests of my project.

I contend that the transparency accounts I consider fail to give a satisfying explanation for why transparently acquired beliefs about one's own doxastic states are epistemically warranted. That is, they do not explain why these beliefs count as knowledge. On my view, the question of how a subject's judgment that *p* warrants her in self-ascribing the belief that *p*, can be answered by considering what is involved in making that judgment, and how precisely it relates to the relevant first order belief.

The positive account I advance draws primarily on work due to McHugh (2011, 2012), O'Brien (2007), and Peacocke (1998, 1999). While the views they put forward are not transparency accounts, the ideas developed by these accounts can play an important role in explaining how transparent self-knowledge is warranted. I will now briefly summarise these accounts.

#### **0.3.4.1 The Reasons Account**

The reasons account was first proposed by Peacocke (1999) and has been defended by McHugh (2012). On this view, a distinction is drawn between *conscious, occurrent judgments* and *standing dispositional beliefs*. According to the reasons account, a subject's conscious judgment that *p* acts as a reason for her to self-attribute the belief that *p*. This is due to the fact that judgments are standardly events of belief formation. Under normal circumstances, if a subject judges that *p* she believes that *p*.

For a subject to go from judging that *p* to the belief that *she believes that p* is matter of making a metaphysical transition, which is rational on the condition that the subject making it is 'conceptually equipped to make the self-ascription' (1999: 214). As Coliva clarifies, the subject would need to be in possession of 'the first person concept, the concept of belief and those concepts which are necessary for the specification of the content of your belief' (2008: 15).

#### **0.3.4.2 Agent's Awareness**

O'Brien develops an account of self-knowledge of actions, based on the idea that agents have a particular kind of non-observational, non-propositional awareness of their own actions. This includes mental actions such as judgment. This awareness is taken to provide non-inferential grounding for self-knowledge.

On the agent's awareness view, mental actions that are a product of the subject's rational assessment give the subject warrant to immediately self-ascribe her thoughts. That is, such awareness provides warrant for knowledge of one's own mental actions without the need for 'mediating acts or representations' (2007: 183-4). So one would not, for instance, need to make any kind of inference in order to go from having agent's awareness of one's act to self-attributing the act. A subject need only ask herself what she is doing.

#### **0.3.4.3 An Alternative Transparency Account**

In developing my own account of transparent doxastic self-knowledge, I begin by presenting an account of the metaphysical and phenomenological nature of judgment. I argue that judgments are exercises of one's agential capacity to make an assessment of a subject matter, and as such are conscious events of which one is phenomenally aware. I claim that the phenomenal awareness one has of one's own judgments involves recognising them as exercises of one's own agency, and thus as expressions of one's epistemic commitment to a given content. I refer to this awareness as 'judgment-mineness'.



I go on to argue that judgment-mineness entails, insofar as it partly constitutes, knowing that *I judge that p*. Because judging is *ceteris paribus* belief formation, knowing that *I judge that p* puts me in a position to knowledgeably self-ascribe the belief that *p*. I argue that we make such self-ascriptions by way of an inference, from the implicit premise *I judge that p*, which is made available to us through judgment, to the conclusion *I believe that p*.

## 0.4 Structure of Discussion

In Part 1 of this thesis (Chapters 1-3) I provide critical analyses of three transparency accounts. In Part 2 (Chapters 4-5) I advance my alternative account.

### *Chapter 1: Moran's Agentialism*

In Chapter 1 I consider the agentialist view advanced by Moran, one of the most significant and influential accounts in the literature. I have chosen to focus on this account in the first instance partly as a means of introducing the idea that the transparency of self-knowledge is importantly connected to the notions of responsibility and commitment.

I argue that, on the reading I present, the agentialist account faces problems due to the fact it is seemingly committed to a version of the 'uniqueness thesis'. That is, it is committed to the thought that a subject will take a given set of reasons to permit just one rational doxastic-response. However, if a single set of reasons can be taken by a subject to support *p* in such a way that permits believing that *p* without requiring it (i.e. permits both belief that *p* and suspension of judgment over *p*), then the transparency procedure described by Moran will not always yield knowledge of attitudes that conform to reasons.

I further argue that the epistemology of transparent self-knowledge as described by Moran is largely unsatisfactory due to its reliance on a form of transcendental reasoning.

### *Chapter 2: Reflectivism*

In Chapter 3 I evaluate Boyle's reflectivist account, which is the view most similar to my own. However, I contend that the account suffers from the fact that it relies upon a problematic metaphysical conception of belief, according to which beliefs are themselves exercises of agency, as opposed to the passive product of mental action.

I further argue that reflectivism makes belief 'over-reflective', in that it requires the subject to not only regard *p* as true, but also to endorse believing that *p*. Precisely what this means is unclear, but understanding belief as constituted by endorsement has potentially serious consequences for

## Introduction

the account. At best it means committing to over-complex accounts of epistemic failings such as self-ignorance and akrasia. At worst it entails a vicious infinite regress.

### *Chapter 3: Inferentialism*

Chapter 3 looks at the inferentialist account of doxastic self-knowledge advanced by Byrne. Through this account we can evaluate the idea of transparency without conceiving of it as connected to rational agency. I conclude that ultimately the inferentialist transparency account fails to give a satisfactory explanation for why transparent self-ascriptions of belief are epistemically warranted.

I advance a series of objections to Byrne's view which develop the dilemma initially posed by Boyle (2011). Through these objections I show that the inference captured on the doxastic schema cannot be rational, and thus cannot be knowledge conducive.

### *Chapter 4: The Metaphysics of Judgment*

In this chapter I begin outlining my positive account of transparent doxastic self-knowledge. The account I advance, like the reasons account, relies on a distinction between conscious occurrent events of judgment and standing belief states. On my account, the key to understanding the warrant for transparent doxastic self-knowledge lies in understanding the distinctive nature of judgment, both metaphysically and phenomenologically speaking, and how it differs from belief.

Accordingly, Chapter 4 details a metaphysical account of judgment, and the belief-judgment distinction. Importantly, I claim that judgments are mental actions. I build a case for this construal of judgment by appeal to the following considerations: (1) Judgments are things for which we hold subjects accountable in such a way that is likely to be grounded by action, and (2) Judgments aim at truth. These features, combined with the fact that judgments are occurrent events, support the judgment-as-action thesis.

### *Chapter 5: Phenomenal Awareness of Judgments*

In Chapter 5, the second chapter in the positive section of the thesis, I give a phenomenological account of judgment. Specifically, I advance the view that when a subject makes a judgment she is necessarily aware of this judgment as *hers*. In being aware of the judgment as hers, she understands it as (1) a product of her own agency, and (2) entailing her epistemic commitment to its contents. This awareness of her own judgments constitutively grounds *tacit self-knowledge of her judgment*. I argue that, since knowledge that *one judges that p* entitles one to knowledge that *one believes that p*, judgments provide epistemic warrant for doxastic self-knowledge.

Finally, I propose an account of how we go about making transparent self-ascriptions of belief. That is, I turn my attention away from the source of warrant for transparent self-knowledge and towards the procedure, or method, by which we transparently attribute beliefs to ourselves. On my account, transparent self-ascriptions are made by way of an inference from *I judge that p* (a premise made available through judgment) to *I believe that p*. My account thus ultimately incorporates elements of both the agency-based accounts and the inferentialist transparency account.



## Part 1: A Critical Analysis of Three Transparency Accounts

### Chapter 1     **Agentialism**

In this chapter I consider the highly influential ‘agentialist’ transparency account advanced by Moran (2001, 2003). On this view, one learns of one’s own beliefs transparently by reflecting upon the reasons relevant to the subject matter of those beliefs, in a process referred to as ‘deliberation’. The fact that this process yields self-knowledge is supposedly explained by rational agency; it is because human subjects form beliefs in response to reasons that considering those reasons alone can itself produce knowledge of one’s own beliefs. Beliefs on this view are expressions of agency insofar as they represent commitments derived from reflection upon one’s evidence.

I contend that Moran’s agentialism makes a number of plausible claims about the nature of belief and its connection to rational agency. Moreover, it succeeds in capturing a distinctively first-personal form of self-knowledge. However, while much of the account is appealing and suggestive, there are fundamental gaps that leave it ultimately unsatisfying. In particular, the account fails to give an adequate explanation of the warrant for transparent doxastic self-knowledge.

This chapter is structured as follows: In section 1.1 I set out my reading of Moran’s agentialist account in detail. In section 1.2 I advance a critical analysis of the view. The first objection I consider in this critical portion of the chapter concerns the view’s apparent commitment to a version of the uniqueness thesis – it seems to suggest that reasons in support of  $p$  will always be regarded by the subject as supporting a single attitude, namely belief that  $p$ . However, it is plausible that a given set of reasons could be taken by a subject to permit both belief that  $p$  and suspension of belief that  $p$  in some cases.<sup>5</sup> In these cases, the transparency procedure may not yield knowledge of rational beliefs. I claim that, as it stands, Moran’s agentialist account is not equipped to deal with this worry.

The next part of my critical discussion focuses on the conception of rational agency Moran advances to explain how transparent self-knowledge is warranted, according to which it is a precondition of rational deliberation that we think of our own beliefs as being shaped by our reasons. It supposedly follows from this that we have a right to the ‘Transcendental assumption’

---

<sup>5</sup> This objection develops a suggestion due to Way (2011).

(2003: 406) that our reasons determine our beliefs. This assumption in turn is supposed to provide an epistemic entitlement to make transparent self-attributions of belief.<sup>6</sup>

I contend that it isn't clear why the possibility of rational deliberation depends on thinkers taking themselves to be rational, insofar as their beliefs conform to their reasons. Moreover, it is also not clear why we should accept a transcendental assumption as the source of our right to transparent self-attributions of belief; a less obscure explanation is that such attributions are warranted because of the nature of reasons based assessments.<sup>7</sup>

## 1.1 The Transparency Method

Moran's notion of transparency is based on the idea that we can come to know what we believe by reflection upon our *reasons* alone. 'Reasons' in this context refers to those things which a subject conceives as normative reasons, normative reasons being facts that count in favour of something. In the doxastic case, a normative reason for believing something is simply a fact that counts in favour of forming a belief with a particular content. For example, the fact that there is frost on the ground is a normative reason for believing *that it is cold outside*. On Moran's view, the relevant kinds of reasons are those an agent *regards as normative*. Such considerations have the status of being facts from the agent's perspective, but this need not be their agent-independent status. These reasons may also fail to genuinely support beliefs that they motivate. The point is that they are regarded as doing so by the agent. I will thus refer to them as 'agent's reasons'.

Acknowledging the distinction between ordinary normative reasons and agent's reasons can aid us in understanding the very specific conception of rationality Moran utilises in his account. This kind of rationality does not depend on there being correspondence between a subject's agent's reasons and the true state of the world. Rather, it depends on her motivating reasons, the reasons for which she believes, *matching up* with her agent's reasons. Or, in other words, it depends on a subject's agent's reasons also being her motivating reasons. Moran's view is that, insofar as we are rational, our beliefs tend to conform to our normative reasons. So, by reflecting upon these reasons in deliberation we can arrive in a position to make accurate self-attributions of belief.

On this account, for an agent to determine her own belief through deliberation she must treat the question of what she believes about *p* as 'equivalent' to the question of what is true of *p*, and she

---

<sup>6</sup> My understanding of Moran's transcendental argument is indebted to Gertler (2011, 2018).

<sup>7</sup> In Chapter 2 I consider Boyle's reflectivist account, which takes this style of approach.

can only establish this equivalence by regarding the question of what she believes 'in a *deliberative spirit*' (2001: 63). That is, she must regard the question of what she believes about  $p$  as a matter of deciding what she is to believe about  $p$  and committing to this stance.<sup>8</sup> Reaching a conclusion regarding the truth of  $p$  provides both an answer to the question of whether  $p$  and to the question of what one believes, since the question 'do I believe  $p$  is true' is 'transparent' to the question 'is  $p$  true?' (2001: 61).

According to agentialism the deliberative procedure is the process by which we attain knowledge of *rational beliefs*. That a belief could have been learned about via the transparency method seems to be a necessary condition for the belief's being rational. That is, a rational belief must meet what Moran calls the 'Transparency condition'.

When there is an attitude of mine that I cannot become aware of through reflection on its *object*, it suggests that the attitude is impervious to ordinary rational considerations relevant to the maintaining or revising of the attitude. [...] [I]f the persistence of the attitude is impervious in this way to considerations of its justification, the attitude itself cannot be a fully rational one and must instead be seen as a kind of fixation. (Moran, 2001: 107-8)

So, when we describe beliefs as rational in this context, we mean that they are rational *insofar as they are shaped by one's reasons*. Therefore, there is a sense in which one meeting the condition that their beliefs conform to their reasons (what we can think of as 'the rationality condition') can be taken to be *locally* rational. We might think someone who believes that the discoloured patches of grass in their garden are caused by aliens is irrational in that their belief contradicts the facts as they stand in the world, or goes against the available evidence, however they may still retain the form of local rationality described on Moran's account if the belief meets the transparency condition. That is, they are locally rational in the relevant sense if their belief conforms to (and is motivated by) their agent's reasons and could have been learned about through reflection on those reasons. So, they would be rational in this way if the presence of discoloured patches of grass was for them a normative reason counting in favour of aliens landing in their garden; the kind of rationality at stake here pertains to how a subject responds to her own reasons.

---

<sup>8</sup> This does not entail voluntarism – an agent 'decides' what she believes insofar as she engages in the activity of evaluating the reasons relevant to the question of what her belief is. But this does not undermine the thought that we are responsible for our beliefs (Moran, 2001: 120). Hieronymi (2008) gives an interesting account of non-voluntaristic doxastic responsibility, arguing that the thing which makes us responsible for our beliefs is the very same thing which prevents them from being voluntary, namely that they 'embody one's take on an object' (2008: 371).

We can draw a parallel between the conception of irrationality relevant to Moran's account and that involved in cases of *akrasia*.<sup>9</sup> When a subject acts akratically she acts in a way contrary to that which she judges to be the best course of action. Even if her judgement is false, the fact that she has made it can be seen to rationally require her to act in accordance with it. In a similar way, according to Moran, if a subject's agent's reasons support  $p$ , she is rationally required to believe that  $p$ . If an agent were to believe that  $p$  when her agent's reasons (which supply the evidence for her belief content) counted in favour of  $\sim p$ , her belief would not meet the transparency condition, in that it would be impossible to acquire knowledge of it using the transparency method, and the belief would therefore not qualify as rational in the relevant sense.

It is important to note that Moran's view aims to capture certain important characteristics specific to self-knowledge, in particular its *authority* and *immediacy*. Traditionally, to speak of self-knowledge as being authoritative refers to the fact that subjects have a special form of access to their own mental states. This supposedly ensures that their self-attributions can be made with a particularly high degree of accuracy. But this is not the kind of authority Moran appeals to. On the agentialist account, a subject has authority over her beliefs because they are *up to her* (Moran, 2001: 134). This relates directly to Moran's conception of us as deliberators who can learn about what we believe by making our minds up about what *to* believe.

Agentialism further construes self-knowledge as having the unique quality of being immediate, primarily in that when an agent makes up her mind to believe  $p$ , she does so *non-inferentially*, and without appealing to evidence in the form of observations about her own behavior.

The "immediacy" of self-consciousness and first-person authority, the fact that I can be aware of my belief without inference or evidence, is a function of the fact that the information about myself that I would gain through inference or evidence about myself is ruled out as irrelevant to the question of what I am to believe. (2001: 134)

As O'Brien puts it, according to Moran 'given the subject's understanding of deliberation, the subject knows that all that needs to be done to find out what she believes is that she reach a conclusion, via deliberation, about what is the case' (O'Brien, 2003: 379). When a subject concludes that  $p$ , she is in a position to know, without the need for inference or further reasoning, that she believes that  $p$ .<sup>10</sup> Knowledge of her belief that  $p$  could also be expressed as being simultaneous with arriving at the conclusion that  $p$ .

---

<sup>9</sup> There is disagreement over whether or not akratic action is possible, but it is helpful for us to use it as a point of comparison.

<sup>10</sup> I consider an inferential transparency account in Chapter 3.



## 1.2 The Agentialist Epistemology

The notion of transparency defended on Moran's initial account is *prima facie* troubling, since it is not clear why a conclusion about a given subject matter should warrant a subject in self-attributing a belief to herself. In other words, the connection between reasons and doxastic attitudes is under-explained. It needs to be made clear why agent's reasons provide grounding for knowledgeable self-attributions of belief. According to Shoemaker, Moran only explains how we get to the affirmation of the 'content proposition', that is, the account only explains how we reach an answer to the question 'is *p* true?' and not why we are entitled to an accompanying answer to the question 'do I believe *p* is true?' (Shoemaker, 2003: 397-398). O'Brien refers to this *prima facie* issue as the 'two topics problem', a term suggested by Moran himself (2007: 103).

Moran's account of the epistemic warrant for transparent self-knowledge is explicated in his response to Shoemaker (Moran, 2003: 405-6). The response begins by considering how transparency could be employed in the specific case of answering the question of whether he believes it is raining:

'I *would* have a right to assume that my reflection on the reasons in favor of rain provided an answer to the question of what my belief about rain is, if I could assume that what my belief here is was something determined by the conclusion of my reflection on those reasons. An assumption of this sort would provide the right sort of link between the two questions.' (Moran 2003: 405)

So, if it is the case that we can *assume* our beliefs are shaped by our rational reflections on the world, we can establish the connection between the question of whether *p* is true and whether I believe that *p*, thus answering the question of how transparent doxastic self-knowledge is warranted. But Moran needs to show where the warrant for this assumption comes from. He does so through a Kantian transcendental argument.<sup>11</sup>

I take the simple Transparency condition to have the surprising consequence that it brings us up to the region of something like a Transcendental assumption of Rational Thought as it has figured elsewhere in Kantian and post-Kantian philosophy, as venerable and familiar as it is obscure. (2003: 406)

---

<sup>11</sup> Transcendental arguments are found throughout Kant's writings, and represent one of his most significant contributions to modern philosophy. For famous examples see *The Critique of Pure Reason* (1781/1787, 1987 ed.: A84–130/B116–169, B274–279). An argument of this kind is also developed by Burge in his account of the epistemic warrant for self-knowledge (1996).

## Agentialism

Standardly, transcendental arguments begin from a widely accepted, seemingly obvious or intuitive premise about the nature of thought or experience. It is then argued that a more controversial premise must be true in order for the first premise to hold. That is, the more controversial idea is taken as a *necessary condition* on the intuitive premise's being true. This might mean the controversial premise is regarded as logically necessary for the first premise, however it could also mean that the controversial idea is the only available means of explaining the first premise. Moran's transcendental argument can be expressed as follows:

P1 We are rational deliberators.

P2 It is a precondition of rational deliberation that we conceive of our beliefs as being shaped by our reasons.

∴

C1 We have a transcendental right to assume that our reasons shape our beliefs.

C2 Because we have a transcendental right to assume our reasons determine our beliefs, we have a right to self-attribute beliefs on the basis of our reasons (i.e. transparently). This explains why transparent self-attributions are warranted such that they count as knowledge.

So, our 'transcendental right' to assume that our reasons shape our beliefs is a consequence of the fact that we must think of ourselves in this way in order to engage in rational deliberation, which we uncontroversially do. This right is supposedly what entitles us to make transparent self-attributions.

It is worth taking a moment to explicitly state the significance of the notion of *entitlement*. We should keep in mind that agentialism does not suppose that whenever a subject deliberates over her reasons she gains self-knowledge. Rather, the point is that deliberation can bring forth a right, similar to that which we find in cases of legal entitlement, to self-knowledge. That is, it gives the subject an epistemic right to make knowledgeable self-attributions of belief, a right which she may or may not exercise. It is important to keep this fact in view, since it shows the account to be less demanding than it might appear.

### 1.3 A Note on Deliberation

Throughout Moran's account, deliberation is treated as centrally important to the acquisition of doxastic self-knowledge. The account suggests that it is through deliberative reflection on one's reasons that one comes to be in a position to transparently self-attribute beliefs, insofar as deliberation involves treating the question of whether *p* as equivalent to the question of whether

one believes that *p*. Deliberation is thus conceived as *the* way in which we acquire transparent self-knowledge. This is reflected in Moran's epistemology, according to which we have a right to the transcendental assumption that our beliefs conform to our reasons, precisely because this is a necessary condition on rational deliberation. Thus to speak of a transparency procedure in Moran's work is to refer to critical reflection upon one's reasons. However, I believe that some of the attractiveness of Moran's view hinges on exactly how we understand the notion of deliberation.

The natural way to think of deliberation is as a presently unfolding process of reasoning. This would presumably involve things like making inferences and weighing different reasons against one another. Essentially, it would be a kind of conscious, reflective activity which took place for a measurable length of time, i.e. a process with a beginning, middle, and end. However, many of our world-directed judgments are made without such a process taking place. Sometimes things are just evident to us in a way that does not require a protracted period of rational reflection. Nonetheless, such judgments are standardly made *in response to reasons*.

In being responses to reasons, even judgments that are non-deliberative – in the sense that they do not involve a drawn out process of reasoning – can be rational expressions of our *agency*, of our capacity to form attitudes in response to (what we regard as) facts. I take it that the best version of Moran's account is one in which we have an entitlement to transparent self-knowledge in virtue of making a *rational, reason-responsive* judgment on a subject matter. But if this is the case, then it can't be that deliberation, conceived as an extended or effortful process, it is required for transparent self-knowledge.

So how can we accommodate seemingly non-deliberative, but nonetheless reason-sensitive judgments in Moran's account? I think the answer lies in adopting a broader understanding of what can qualify as deliberation in the relevant sense. If we think of deliberation as referring to the event of rationally evaluating one's reasons, then we can allow this to encompass both reflection upon one's reasons that happens over time, and involves things like inferential reasoning and evidence weighing, as well as reasons-sensitive evaluations which happen *at a time*.

If we re-interpret the deliberative procedure as not so much a process of reflecting upon reasons, but an event of rationally *responding* to reasons, which can happen either at a time or over time, then we can reconstruct Moran's account to accommodate the idea that judgments reached without a temporally prior process of conscious, reflective reasoning can nonetheless give us a right to transparent self-knowledge. On this picture, it is preferable to employ Gertler's terminology and refer to the view as describing a transparency 'method' rather than a procedure.

While ‘procedure’ is suggestive of a process that happens over time, method is temporally neutral.<sup>12</sup>

To be clear, I do not presume that Moran would endorse the conception of deliberation I have proposed here. Indeed, as I discuss in 1.4.2, there is reason to suppose Moran has a more demanding account in mind. However, for now I want to draw attention to the fact that rational, reasons responsive assessments need not be preceded by a drawn out process of deliberation in order to be conceived as exercises of a subject’s capacity of rational agency. For this reason it is *prima facie* plausible that non-deliberative judgments have the special epistemic feature of authority, and as such can provide grounds for knowledgeable self-ascriptions of belief.

### 1.4 Objections to Agentialism

In this section I advance critical responses to Moran’s agentialist account. In section 1.4.1 I develop an objection which explores the idea that some beliefs which meet the rationality condition (i.e. conform to a subject’s agent’s reasons) might not meet the transparency condition. This objection appeals to the idea that the account relies on a version of the ‘uniqueness’ thesis, according to which a subject cannot regard her agent’s reasons as supporting more than one rational attitudinal response. I contend that a single set of agent’s reasons in support of *p* can plausibly be taken by the subject to support both belief that *p* and suspension of belief that *p*. In cases where this happens, the transparency procedure may not yield self-knowledge. I claim that, as it stands, Moran’s view is not equipped to deal with this worry.

In section 1.4.2 I turn my attention to the epistemology of Moran’s agentialism. I argue that it isn’t clear that the possibility of rational deliberation depends on a subject taking herself to be rational, insofar as her beliefs conform to her reasons. I further claim that asserting a right to a transcendental assumption about the relation between reasons and beliefs does not tell us how a subject goes from making a judgment on the basis of reasons to making a self-attribution of a belief with the same content. So, there are seemingly two key questions, regarding the agentialist epistemology, for which the account does not have a satisfying answer: 1) why one must conceive of oneself as rational (in that one’s beliefs conform to one’s reasons) in order to engage in rational thought, and 2) How one transitions from a judgment about *p* to believing that *one believes that p*.

---

<sup>12</sup> Insofar as ‘method’ refers to the way you do something, or the technique for accomplishing a task, it does not imply anything about the temporal character of an event.

The critical points I present here are not intended as knock-down objections to agentialism, rather they aim to bring out the ways in which Moran's suggestive account is lacking specific details that would make it a full and coherent theory of self-knowledge. In the next two chapters I consider two very different approaches to filling in these details: Boyle's reflectivist account, which preserves Moran's idea that doxastic self-knowledge is importantly connected to rational agency, and Byrne's inferentialism, which advances a more minimalist picture on which the transparency of doxastic self-knowledge is explained without appeal to rational agency.<sup>13</sup>

#### 1.4.1 The Uniqueness Objection

A potential problem with transparency, conceived as a means by which we can arrive at accurate self-ascriptions of belief, is that fact that our beliefs sometimes come apart from our reasons, and the judgements made through reflection on these reasons. It might be the case that I *believe* Sanders will be president, but it does not seem to *necessarily* follow from this that my deliberation over the question of whether this will happen will yield the judgment-conclusion that *Sanders will be president*. However, Moran's account is not so demanding as to assert that we must be perfectly rational, and he fully accepts that it is possible for us to hold beliefs which do not conform to our reasons. As he writes;

A fairly modest version of "first-person authority" will understand it not as entailing either infallibility or perfect access, but as a feature of discourse, as the authority a speaker is granted to declare his thought and feeling, and have that declaration count (normally decisively) as telling us what the person's attitude is. (2001:121)

Nonetheless, a problem seems to arise for Moran if the transparency method cannot yield knowledge of beliefs which it *should* be able to according to his view. On my reading of Moran, wherever a subject's belief meets the condition of being *rational* in the relevant sense (in that it conforms to an agent's evidence), the transparency condition (that the subject can learn of the belief by reflecting on her reasons) must also be met. I will demonstrate that this part of the account is reliant on the controversial idea that a deliberating subject cannot regard her reasons as rationally supporting more than one attitude response. In this way it relies on a version of the 'uniqueness' thesis.

I intend to argue against this form of uniqueness by considering cases in which subjects could plausibly regard their reasons as permitting more than one attitude response. If this is possible, then a subject may possess beliefs which fulfil the *rationality condition*, in that they conform to

---

<sup>13</sup> Or at least not the rich form of rational agency Moran and Boyle appeal to.

her reasons, but which fail to meet the *transparency condition*, since they could not be learned about through the transparency method.<sup>14</sup>

#### 1.4.1.1 Agent-Relative Uniqueness

The uniqueness thesis may be construed in a number of ways, but can be broadly defined as follows: *from any given set of evidence there is one uniquely rational conclusion to be drawn*.<sup>15</sup> In other words, a single set of evidence cannot permit more than one rational response. I intend to show that Moran's account is seemingly committed to a particular form of the uniqueness thesis. I will argue that this version of the uniqueness thesis is implausible, and in doing so defend a particular form of permissivism, i.e. *the thought that there can be more than one rational response to a body of evidence*.<sup>16</sup>

For a subject to be able to use Moran's transparency method to learn of her own doxastic attitude towards *p*, it must be that she takes her reasons to permit a single rational attitude. If she were to regard her reasons as permitting more than one attitude, then reflection on these reasons alone would not be conducive to self-knowledge.

The form of uniqueness relevant to Moran's view can thus be captured as follows:

*A subject who engages in rational deliberation over p can only regard her reasons as permitting a single attitude response towards p.*

Note that the above form of uniqueness is an *agent-relative* one. That is, it depends not on the normative status of a body of evidence (in this case a set of agent's reasons) but on how the *subject* treats this evidence. Whether it is possible for a body of evidence to permit more than one rational conclusion is essentially irrelevant. What is important is whether a subject can regard the evidence as supporting more than one rational response.

I will assume that the conclusion of a subject's deliberation over *p* can be characterised by one of the following classificatory concepts: acceptance, rejection, or suspension of judgement. I take these to normatively support the following corresponding doxastic attitudes: belief, disbelief, or suspension of belief.<sup>17</sup> When an agent concludes deliberation over the question of whether *p* is

---

<sup>14</sup> Often, Moran presents deliberation as a matter of deciding what to believe. In this way, it is conceived as both a method for acquiring doxastic self-knowledge *and* an event of belief formation. However, in spite of this presentational convention, the transparency method is also meant to yield knowledge of previously acquired doxastic attitudes. This is where my attention is presently focused.

<sup>15</sup> See Kopac and Titelbaum (2016) for a detailed overview of the debate around uniqueness.

<sup>16</sup> Kelly (2013) puts forward one of the leading arguments for permissivism.

<sup>17</sup> Feldman (2007) likewise takes these to be the three possible doxastic states. The doxastic status of 'suspension of belief' is however a point of contention. While I am keen to assert that suspending

true, she will accept that  $p$ , reject that  $p$ , or suspend making any judgement at all on the matter, and in doing so regard her reasons as supporting belief that  $p$ , belief that  $\sim p$ , or suspension of belief regarding  $p$ . That is to say, the deliberating subject is concerned not merely with answering the question of whether  $p$ , but also with the second-order question which doxastic attitude her reasons support. A single event of deliberation, i.e. reflection upon one's reasons, yields answers to both of these questions.<sup>18</sup>

But if it is possible that a subject could regard her reasons as permitting two incompatible responses, i.e. believing that  $p$ , and suspending judgement on  $p$ , then reflection upon her reasons alone would not be sufficient for learning what her attitude is. This is problematic because it suggests that the transparency method will in some cases lack power: if it is the case that the evidence for the content of a belief state could count equally in favour of two different rational responses, deliberation will in this instance be unable to generate an answer to the question of which doxastic attitude the subject possesses.

Way (2007) develops a criticism of transparency based on the idea of uniqueness. For the most part his discussion is focused on why the transparency method is not a reliable means of coming to know one's *intentions* (2007: 227). He asserts that we are often faced with having to choose between two or more options for which we have equal reasons in favour of each, and draws on an example described by Bratman (1987: 23). Bratman imagines he has a choice between taking either route 101 or route 208 to San Francisco. In the example, both are equally attractive and effective options. Thus the intention to take either would be a rational attitude.

Way supposes that Bratman decides to take the 101, and in doing so forms the intention to take this road. The knowledge he possesses of this intention has the qualities which we generally take to be distinctive to first-personal self-knowledge, in that it is both *authoritative* and *immediate*. However, it cannot be the case that Bratman could come to learn about his intention by reflecting upon the question of whether to take the 101, since the reasons relevant to this question equally support both taking the 101 and taking the 208.

Way concludes that cases such as Bratman's, which arise in practical reasoning regarding how one should act, demonstrate that uniqueness does not hold in the case of intentions, and therefore show that the transparency method is not a reliable process by which we can acquire knowledge

---

judgement on a matter is more than merely lacking a doxastic attitude, if it is a class of attitude in itself it is not clear exactly what sort of attitude it would be.

<sup>18</sup> I will discuss this idea, and why it is important to the account, in more detail in the objections section of this chapter.

of these attitudes.<sup>19</sup> However, as Way notes (2007: 228), whether it is reliable in cases of theoretical reasoning regarding beliefs is a further, altogether more controversial question, and it is this issue which I now intend to explore.

I aim to demonstrate why we should be skeptical about the claim that the agent-relative form of uniqueness applies to beliefs, and consequently skeptical about the transparency method as a reliable process through which we can learn about rational beliefs. I will develop this suggestion by appeal to examples of the kinds of cases in which one is likely to regard one's reasons as not obviously supporting a single rational conclusion. In doing so, I will attempt to flesh out Way's proposal that there may be evidential states which rationally permit belief in *p* without requiring it, thus permitting both belief that *p* and the *suspension of belief* that *p* (2007: 228).

Testimony cases are one of the most promising places to look for instances of theoretical reasoning in which subjects plausibly regard their reasons as having more than one rationally permissible conclusion. In many of these cases, where the source of information can safely be assumed to be neither especially unreliable nor especially authoritative, the fact that a subject takes her testimony derived reasons to support *p*, and therefore to support belief that *p*, would not entail that the subject regards herself as rationally *obligated* to believe that *p*. So, it seems that theoretical reasoning on the basis of testimony can often have more than one rationally permissible attitude response from the subject's perspective.

The following example is intended to demonstrate the way in which subjects often treat evidence (i.e. agent's reasons) derived from testimony:

Jack is the caretaker of a hotel where he lives with his wife Wendy and their son Danny. One day Jack is looking for Danny. He asks Wendy if she has seen him. Wendy tells Jack that she has, and that Danny is in room 237. Jack has no reason to suppose that Wendy is an unreliable source of information in this case, however he also knows that it is at least possible that Danny has left room 237 since Wendy last saw him there.

Let us assume for the sake of argument that Jack in fact forms the belief that Danny is in room 237, on the basis of Wendy's testimony. However, when Jack reflects upon the reasons relevant to the question of where Danny is, which he derived from Wendy's testimony, he finds them to rationally permit *both* believing that Danny is in room 237 *and* suspending judgment on the matter.

---

<sup>19</sup> The transparency method would be unreliable in the sense that it would not be possible to generate a single, uniquely rational answer in certain cases, as opposed to unreliable in that it would produce inaccurate results.



The crucial point is that Jack merely reflecting on his reasons will not yield knowledge of what his attitude is. That is, such a method for discovering what he believes will not, in this case, have the power to produce an answer. Recall that whether one of these options is *in fact* more rational is not relevant for our purposes. Nor is it important what leads Jack to believe rather than suspend belief.<sup>20</sup> The key point is that Jack could plausibly regard them as equally permissible. This being the case, Moran's transparency method as I have construed it could not yield knowledge of Jack's belief in this instance, since it would not be possible for him to determine what his attitude is solely by reflecting on his reasons. This is in spite of the fact that he has a belief which conforms to his reasons.

My suggestion should carry some appeal insofar as it expresses the broadly uncontroversial thought that believing on the basis of testimony is something which, by its nature, involves uncertainty, in that there is always the possibility of being misled (maliciously or not). Though there is much disagreement within the literature on testimony<sup>21</sup>, it is generally accepted across the board that risk, and an assumption of reliability/sincerity on the part of the speaker, is inherent in the act of accepting testimonial evidence.

Interestingly, Moran's own account of testimony (2005), which focuses on the relationship between the speaker and the subject being addressed, is consistent with my argument. He draws a distinction between believing *the person* providing the testimony and believing *the proposition* expressed by the testimony, the latter of which follows from the former. His claim is that, in standard cases, 'when the hearer believes the speaker, he not only believes what is said but does so on the basis of taking the speaker's word for it' (2005: 2).

Clearly, Moran would have to accept that risk is an unavoidable cost of believing based on testimony – to 'take the speaker's word for it' just *is* to accept their testimony in spite of the fact that it might be false. Taking someone at their word would not be something one is rationally *required* to do in most testimony cases, rather it is something one is rationally permitted to do. It seems reasonable to think on this basis that subjects could regard their testimony-derived reasons as permitting more than one attitude response.

---

<sup>20</sup> For what it's worth, I take it that subjects will ordinarily be more likely to suspend judgment in high risk cases, i.e. in cases where the cost of believing falsely is especially great. What I have claimed here does *not* entail that in cases where a subject determines her reasons to support more than one attitude response she can choose which attitude to take.

<sup>21</sup> One of the major debates in the literature on testimony arises between Humean reductionist accounts and non-reductionist accounts.

There may be cases other than those of testimony which would also not rationally require a single unique conclusion. To demonstrate, I will consider a similar example to the testimony case discussed above:

My brother, Toby, is travelling around Europe. When attempting to determine what I believe about his current location, I note that this morning I received a postcard from Berlin, and saw a photo of the Brandenburg Gate on his Instagram. During my deliberation it also occurs to me that even if he had been in Berlin, he could by now have left the city. Nonetheless it still seems likely that this is his current location. I interpret my reasons as permitting me equally to (t) believe that Toby is in Berlin *or* (s) to suspend this belief, i.e. I take myself to be permitted but not *required* to believe that *Toby is in Berlin*. Deliberation seemingly results in me regarding it as acceptable to believe that Toby is in Berlin, without thinking I *must* believe that *Toby is in Berlin*, since both (t) and (s) would, according to my own reasoning, be permissible, *rational* conclusions to the same reasoning process.

Moreover, the above responses would be rational in the sense that they would meet the rationality condition. This has troubling implications for Moran, since although I may be in a position to know which of these conclusions I had affirmed, it would *not* be possible to do so using the transparency method. Deliberating about whether my reasons suggest that Toby is in Berlin will only get me as far as establishing two possible judgment-conclusions. Further investigation, or the employment of other reasoning, would be required to learn of my own attitude towards my brother's location.

To suggest that *in deliberation* there is but a single conclusion a subject could take to be rationally acceptable to the question of whether *p* is, at the very least, problematic. It is plausible to suppose there are cases such as the ones I have described in which a subject could take herself to be equally justified, on the basis of her reasons, in either forming a belief or suspending belief, and that in such cases merely reflecting on one's reasons could not *in itself* tell the agent which of these she has done. Of course, the mere fact that Moran's view appears committed to a controversial thesis does need not lead one to dismiss it outright, but it is nonetheless a problem for Moran that his view, if it is committed to uniqueness, hinges on an exceptionally strong claim. It is consequently a far more demanding account than it first appears, and presumably a significantly more demanding account than Moran himself wishes to give.

Allow me to here reiterate that the uniqueness thesis in a broad sense need not be proven false in order for Moran's commitment to it to be problematic. The key point for our purposes is that even if the uniqueness thesis is correct from a normative stance, if it can be the case that agents rationally *take* their reasons to equally permit two different rational conclusions, then the

problem still stands for Moran that the transparency method will not in such cases lead to knowledge of one's own belief. So, Moran's account is committed to a psychological claim as opposed to a normative one – if the transparency method is to generate an answer to the question of which attitude one holds, it must be the case that agents can, in regard to any question of deliberation, only rationally take themselves to have reasons permitting a single doxastic attitude. In other words, *he assumes that rational agents are in their own deliberations committed to uniqueness*, and as I hope to have demonstrated it is far from obvious that this is the case.

#### 1.4.1.2 A Response to the Uniqueness Problem

One possible way of responding to the uniqueness objection is by denying that in deliberation a subject is engaged in second-order reflection upon her reasons. That is, to reject that the subject is thinking about what her evidence supports, and instead take her to be reflecting directly upon the relevant subject matter.<sup>22</sup> On this understanding, the subject is not thinking about whether her reasons support  $p$ , but simply whether  $p$ . In this sense, she attends directly to the question of whether  $p$ . This would serve as a response to the worry I raise, as the uniqueness objection crucially relies on the conclusion to the subject's deliberation involving a decision about her reasons.

To clarify the above thought, if we regard deliberation as *merely* a matter of making a judgment-conclusion about  $p$ , and not as a matter of determining which attitude response one's reasons favour, then deliberation which does not conclude with the subject either accepting  $p$ , rejecting  $p$ , or suspending judgment on  $p$ , is indicative only of the fact that she lacks a reason-responsive doxastic attitude towards  $p$ . If a subject with the reason-responsive belief that  $p$  only considers whether  $p$  in deliberation, then she will conclude that  $p$ .

##### 1.4.1.2.1 Counter Response

The above version of the view brings forth its own problems. If we deny that deliberation on the agentialist view involves a form of second-order assessment of one's reasons then the role given to reasons, and the transcendental assumption that is supposed to explain the epistemic warrant for transparent self-knowledge, seems undermined.

It is claimed that we have a right to assume our reasons shape our beliefs, since this is a precondition of rational deliberation. This right is in turn supposed to give us the further right, or

---

<sup>22</sup> Thanks are due to Brie Gertler for raising this objection in correspondence.

entitlement, to make transparent self-attributions of belief. The transcendental argument therefore suggests that transparent self-knowledge is the product reflecting upon one's reasons, and treating these as indicators of our doxastic states. If deliberation is a matter of evaluating evidence in the form of reasons, it seems to follow that the conclusion of deliberation will be, at least in part, one about that evidence. If this were not the case, it would be unclear why Moran's view appeals to reasons at all. That is, if all the subject is doing is evaluating a subject-matter, the reasons talk in this account becomes redundant.

Undermining the role of reasons in the manner described would have a serious consequence for the account, in that it would prevent it from having the epistemological structure Moran describes. If a deliberating subject were only concerned with evaluating the subject-matter  $p$ , as opposed to evaluating the evidence pertaining to  $p$ , it would have to be the case that the subject was in a position to self-attribute beliefs knowledgeably purely on the basis of some fact about the world. But this leaves the source of warrant for transparent self-attributions mysterious. One cannot appeal to the transcendental assumption, because this relies on the subject conceiving of the conclusion of her deliberation as a conclusion about her reasons, and not about  $p$  itself.

Moran's agentialism would thus end up looking similar to Byrne's inferentialism, according to which doxastic self-knowledge is grounded by a fact about the world (though Moran would deny this grounding is inferential). More importantly, agentialism would face similar problems to those facing Byrne's account. One of the major problems for inferentialism is that it is unclear how it could be rational for one to reason from some fact about the world,  $p$ , to *I believe that  $p$* .<sup>23</sup> The problem for the revised interpretation of Moran's account is that a subject would have no right to assume that her beliefs ordinarily conform to the state of the world, partly because this is not a necessary condition for rational deliberation, but also because such an assumption could not be rational. Thus it could not be the case that she reflects solely upon the state of the world to acquire knowledge of what she believes.

So it seems Moran's agentialism faces a dilemma. Either the uniqueness objection applies, or else central aspects of the view, including the role of reasons, become idle. In the latter case, the account cannot fulfil the second desideratum for a successful account of self-knowledge (outlined in the introduction to this thesis), which states that such an account must capture how knowledgeable self-attributions of belief are warranted.

The dilemma facing Moran's agentialism arises not from inherent inconsistencies with the view itself, but rather from the fact that it simply not clear how key aspects of the account work. If

---

<sup>23</sup> See 3.4 for a detailed critical discussion of inferentialism.

Moran's view is that deliberation concludes with belief formation, or in other words that belief formation is somehow constitutive of deliberation, then there may be a route to addressing the uniqueness worry. However, as it stands, this is not something the account is equipped to explain.

#### 1.4.2 The Role of Rational Agency

So far, I have argued that, on the account Moran advances, a subject may not always be able to learn about beliefs that conform to her reasons by reflecting upon those reasons alone. Thus beliefs that satisfy the *reasons condition* may not satisfy the *transparency condition*. I considered whether we can solve this by understanding deliberation as a merely first order evaluation of a given subject matter, as opposed to also being a second-order evaluation of reasons pertaining to that subject matter. But this seemingly undermines the role of reasons on this account, and in doing so renders the transcendental explanation of our warrant for transparent self-knowledge ineffective.

In this section I consider a problem with the transcendental epistemology itself. The worry I raise stems from its reliance on a demanding conception of rational agency. As a reminder, the crucial premise in the agentialist transcendental argument is the following (which I previously labelled P2): *it is a necessary condition for rational deliberation that one takes one's reasons to shape one's beliefs*. It supposedly follows from this premise that we have a right to assume that our reasons shape our beliefs, and it is this assumption that purportedly entitles us to make transparent self-attributions of belief. However, it is not clear that rational deliberation, i.e. evaluative reflection upon one's reasons, must involve assuming these reasons to shape one's beliefs. Indeed, this seems a demanding condition on rational thought.

Recall that what we are doing in deliberation is reflecting upon reasons relevant to a particular subject matter, with the aim of establishing what this evidence supports. Yet we can think about our reasons, and consequently form beliefs on the basis of these reasons, without assuming that our beliefs standardly conform to our reasons. This is not to deny that there is an important connection between belief and reasons. It is true that rational agents form beliefs in response to reasons. But the transcendental epistemology relies not merely on agents having beliefs shaped by reasons, but on agents assuming that this is the case.

To see why the account appears too demanding, we can consider an example:

Cody comes home to the sound of a recorder. He takes this fact as evidence for his brother Ezra being home. Cody comes to believe that Ezra is home as a result of his reasons assessment.

Do we have to think that, in order to reason in this way, Cody must assume that his reasons shape his beliefs? There seems to be no obvious reason for accepting this, unless we understand Cody's reasoning as inherently self-aware.

We might be inclined to accept the necessity of the transcendental assumption if we take deliberation to be a matter not just of reflecting upon and evaluating one's reasons, but as a matter of doing so *knowingly*, with the *aim of coming to know one's attitude towards a subject-matter*. If a subject regards deliberation as a means of acquiring knowledge of her own beliefs, then it *does* seem necessary for her to assume that her reasons shape her beliefs. If she did not, there would be no explanation for why she should regard reflecting upon her reasons a means of learning about her beliefs.

Alternatively, we could understand Moran's account as making a weaker claim. On a different reading, we might regard the account as claiming that you are entitled to make self-ascriptions transparently insofar as you are capable of reflecting upon the nature of belief and deliberation. On this construal, the subject need not presently regard her beliefs as conforming to her reasons, nor take deliberation to be a way of forming and learning of her beliefs. Rather, she need only have the capacity to reflectively think of her attitudes in this way, and this capacity can be taken to explain why she is warranted in self-ascribing belief through deliberation.

However, the above version of the account raises other issues. For one, we might question how the agentialist transparency method could be rational if the subject merely possessed the capacity to reflectively understand the concepts of belief and deliberation. Moreover, such a version of the account does not square with Moran's expression of the view, whereas the idea that deliberation is an inherently self-aware phenomenon, which involves having the aim of discovering one's own attitude, *is* supported by Moran's discussion:

One must see one's deliberation as the expression and development of one's belief and will... (2001: 94)

In the cases I'm concerned with, the goal of deliberation is to answer a question of the form 'What do I think about X?', that is, a question about one's state of mind. (2004: 413)<sup>24</sup>

---

<sup>24</sup> Note that this quotation suggests deliberating subjects on Moran's view are concerned not only with which attitude is supported by their reasons, as I previously argued, but with which attitude they in fact have.

So, we should understand deliberation on Moran's account as having the aim of self-knowledge. Thus, a subject would only be entitled to make transparent self-attributions of belief on the basis of deliberation when she reflects upon her reasons with the express aim of coming to know her own doxastic states, and an understanding of deliberation as the way to do this. In order to *rationally* understand her own deliberation in this way, she would need to make the assumption that her beliefs conform to her reasons.

However, an account on which entitlement to self-knowledge depends upon the subject understanding deliberation in this way seems excessively demanding. O'Brien expresses a thought along these lines. She suggests that even a subject who had only a partial grasp of the concepts of belief and deliberation, such as a young child, would seem to be in a position to immediately and authoritatively know her own beliefs. She claims that such a subject would be aware of her beliefs in the same way that a more conceptually sophisticated agent is, but would lack the ability to articulate her grounds for this awareness (2003: 380).

I am less confident that a subject who lacked the full concept of belief would be in a position to make knowledgeable self-ascriptions of belief by engaging in deliberation. Certainly, it seems unlikely that such a subject could gain explicit propositional knowledge of her own beliefs in this way, even if she were in a position to have some form of non-propositional awareness of her own judgments. Nonetheless, I want to defend a similar thought to O'Brien's.

A subject need not think of her deliberation as a process by which she forms and can come to know her beliefs in order for her world-directed assessments to entitle her to knowledge of her own beliefs. If we accept Moran's claim that one forms beliefs by actively assessing one's reasons, and I think we should accept that this is a way in which one comes to have beliefs, then we have good reason to think that belief formation is in some sense self-aware. That is, if belief-forming is conceived as an active exercise of one's agential capacities, then it stands to reason that one is aware of what one is doing in such cases.

However, the nature of a subject's awareness of her own rational assessments need not be conceived in the conceptually sophisticated manner in which Moran understands it. A subject who understands her assessments as *her own* is seemingly in a position to understand herself as committed to the conclusions of these assessments. This would in turn seem to suggest that she is entitled to make self-ascriptions of belief when she makes world-directed assessments of her evidence. However, note that it does *not* appear to be the case that the subject must think of her own assessments as the means by which she forms beliefs, and importantly, she need not be explicitly concerned with determining what she believes, in order to have warrant for a knowledgeable self-ascription of belief.

In summary, Moran's suggestion that we are aware of our own rational, active assessments as expressing something about our epistemic commitments is highly plausible. It is further plausible that this awareness puts us in a position to self-ascribe beliefs. However, the agentialist gives an unnecessarily demanding account of the rational conditions under which one is warranted in making a knowledgeable self-ascription.

In the second half of this thesis (chapters 4 and 5) I propose a more detailed account of the precise nature of a subject's awareness of her own active rational assessments. For now I simply note that a subject who assesses her evidence and concludes that *it is snowing* is entitled to self-ascribe the belief that *it is snowing* insofar as she understands the assessment she makes as her own, and *not* only insofar as she conceives of her deliberation as a means of acquiring self-knowledge.

### 1.5 Conclusion

Moran's agentialist account has much going for it. More traditional accounts of self-knowledge understand a subject's authority over her own beliefs as a matter of her having especially direct or secure access to these states. Agentialism recognises that a subject has authority over her beliefs in virtue of the fact that she is agentially responsible for bringing them about. That is, the agentialist account rightly identifies an important connection between a subject's capacity for active, rational reflection upon her reasons, and the formation of her standing commitments.

The account also makes the plausible suggestion that a subject is aware of her rational reasons assessment as being *hers* in some sense. In this way, the subject conceives of her world-directed assessments as representing her own commitments, rather than merely facts about the world. On such a picture, it is likely that the subject's active assessment of *p* entitles her to self-knowledge of her belief that *p*.

However, while the agentialist account gestures towards an attractive account of our warrant for transparent doxastic self-knowledge, it fails to provide a fully fleshed out epistemology. This is partly due to its failure to explain exactly how deliberation, i.e. rational reflection upon one's reasons, grounds belief. As a result of this, the view is left vulnerable to objections such as the uniqueness worry. Moreover, the agentialist epistemology relies on an overly sophisticated form of rational agency. In order for one to have warrant for transparent self-knowledge, one must not only make the assumption that her beliefs conform to her reasons, it seems she must also understand her deliberation as both a means of forming beliefs, *and* of acquiring knowledge of belief states.



The major issue for Moran's agentialism is that it fails to satisfyingly spell out exactly how knowledgeable, transparent self-ascriptions are warranted. That is, it does not give a plausible account of the positive epistemic conditions for transparent self-knowledge. The account appeals to demanding rational conditions for knowledgeable self-ascriptions, and states that such self-ascriptions are made *non-inferentially*. But it does not explain in positive terms exactly how my deliberative assessments provide grounds for doxastic self-knowledge, nor how I go about making knowledgeable self-ascriptions on the basis of these assessments.

The primary things to take away from this chapter are thus as follows: Moran's agentialism identifies an important sense in which we have authority over our own beliefs. Namely, in that we produce beliefs by rationally assessing our reasons. However, the account fails to satisfyingly explain how this fact about the nature of belief formation informs the epistemology of doxastic self-knowledge. The account gets close to a convincing explanation by suggesting that our active, rational assessments are in some way self-aware. But pursuing this line of thought will mean giving a clear statement of the nature of this awareness, *without* relying on an overly demanding account of rational agency. It will also require us to show exactly how the nature of our rational assessments gives us warrant for knowledgeable self-ascriptions. Furthermore, we need to explain how self-ascriptions are made (i.e. whether they are made inferentially, and if not, how else they might come about).

In the next chapter, I consider Boyle's reflectivist account. This view builds upon an alternative reading of Moran's view from the one I have proposed here. I will consider whether this transparency account can convincingly explain the epistemology of transparent doxastic self-knowledge, thus fulfilling the desiderata I initially set out: to capture the distinctiveness of self-knowledge, and to explain why distinctive self-ascriptions are warranted such that they count as knowledge.

Given that reflectivism, much like Moran's agentialism, seeks to explain the warrant for transparent self-ascriptions via our capacity for rational agency, I will also consider whether Boyle's account is equipped to explain the things left mysterious in Moran's account: the nature of our awareness of our own rational assessments, how this awareness warrants knowledgeable self-ascriptions, and how we go about making self-ascriptions. I will argue that while Boyle's account *does* offer explanations for these things, these explanations give rise to new problems.

## Chapter 2 Reflectivism

In this chapter I critically discuss the transparency account proposed by Boyle (2011). The view is largely based upon Boyle's own interpretation of Moran's often ambiguous account of doxastic self-knowledge. According to the account Boyle develops, transparent self-knowledge is possible as a result of the particular metaphysical nature of belief, which is itself importantly linked to the human capacity for rational agency.

Specifically, the transparency of doxastic self-knowledge is taken to be explained by the fact that believing that  $p$  and believing *that I believe* that  $p$  constitute a single 'psychological condition' (2011:235). Supposedly, this is because to believe that  $p$  is to be in an *active* state of knowingly assessing  $p$  as true. Making a conscious self-attribution of belief is thus taken to be a matter of *reflecting* upon one's knowing assessment in the right kind of way. For this reason, the view is referred to as 'reflectivism'.

At this stage, many aspects of Boyle's account are likely to seem obscure. In the first section of the following discussion I elucidate the account briefly sketched above in greater detail. I then develop objections to the reflectivist account.

### 2.1 The Reflectivist Account

As stated, the reflectivist account rests on a particular understanding of the metaphysical status of belief. This metaphysical picture is dependent on belief being understood as an exercise of rational agency. Boyle's unusual metaphysical view of belief is arguably the most crucial part of the account, but it is also one of the most difficult to understand on an initial reading. This is partly because it is seemingly dependent on background commitments that are not always easily visible in Boyle's presentation of the view. In the following sub-section I attempt to clarify what the relevant metaphysical view is.

#### 2.1.1 Belief as an Active State

According to Boyle, belief in human agents is normally a kind of *active state*, insofar as the belief that  $p$  is *constituted* by a sustained commitment to  $p$  being true. In this way, belief is construed as itself an act, or exercise of agency (2011, 2009a). This view stands in contrast to the more widely accepted view that belief is a kind of non-agential state. On more standard views, belief is often regarded as being brought about *by* an exercise of agency (usually understood as judgment) but

belief itself is taken not to involve the active exercising of agential capacities.<sup>25</sup> In fact, on these views, it is usually assumed that belief *cannot* be an exercise of agency, since acts are characterised by the fact that they are ‘occurrent’ events, which beliefs are not. Boyle defines this term as follows:

[The term ‘occurrent’] expresses a conception of the temporal character that is required of something present to consciousness. This term seems apt on the assumption that a conscious mental phenomenon must be something ongoing, something that can be continuous or discontinuous, that lasts a certain definite length of time, that has the sort of duration you could measure with a stopwatch. (2009a: 120-1, n. 5)

The above is a fairly helpful characterisation of what it means for something to be occurrent, however it would benefit from some minor revision. While it seems right to say that the term expresses the temporal character of a conscious episode<sup>26</sup>, it need not be the case that a conscious event is something *ongoing*, with a duration one could measure. It may simply be the case that a conscious mental event occurs *at a time*, rather than *over time*. In this way, though occurrent events are temporally marked, or dated, they need not be conceived as unfolding for a measurable length of time. In this way, a conscious episode need not be an ongoing *process*.<sup>27</sup>

So belief, being a state rather than a dated event, does not look like an exercise of agency, under the standard assumption that acts are occurrent events. A state is not something that happens at a time or over time, in the way that an action like, say, waving my arm, does. It can instead be thought of as describing the way something is: it captures a condition something is in. In this sense, to say one has a belief state would be to make a claim about the way one’s mind is. Thus beliefs are ordinarily regarded as passive rather than active.

Setiya expresses the above point, that beliefs cannot be actions due to their metaphysical character, by appealing to the conventions of grammar:

[...] a metaphysical contrast that is essential to action theory [is] between states, like being tall, and things that can be finished or completed and in that sense done. This

---

<sup>25</sup> Notable proponents of this kind of view include McHugh (2011), Shoemaker (2009a), Peacocke (2007, 1998, 1999), Shah and Velleman (2005). Setiya (2008) takes the view that belief is a non-agential standing state, but is sceptical about the idea of judgment as an exercise of agency.

<sup>26</sup> I take it that all conscious episodes are occurrent, insofar as they are dated events. However I regard the term itself as merely capturing some temporal character, which may be employed to describe subject-independent events (e.g. a tree falling in a forest). Thus on my understanding an occurrent event need not be conscious.

<sup>27</sup> In my positive account (Chapters 4-5) I suggest that judgment is an occurrent event that happens *at a time*, as opposed to a process that unfolds over time.

distinction corresponds to the grammatical notion of perfective aspect. States cannot be, so to speak, perfectly instantiated; they cannot be done. To say that someone was tall, or believed that *p*, is not to say that they completed a performance of being tall, or believing that *p*, as one might complete a performance of walking and thus have walked. It is merely to describe their prior and perhaps enduring condition. [...]The distinction is exhaustive: what can be instantiated by an object can be instantiated perfectly, like walking, digesting, and growing; or it is a state, like believing, desiring, and being tall. (Setiya, 2008: 38)

In holding that belief is both a non-occurrent state and an act, Boyle's view rejects the standard assumption outlined above. We can understand his position more precisely by considering three important distinctions at play in the present discussion. I have drawn distinctions between the following: active and passive, occurrent and non-occurrent, and event and state.<sup>28</sup> Ordinarily, these distinctions are taken to correspond: something can be an *active occurrent event*, or it can be a *passive non-occurrent state*. Hence why, on an ordinary metaphysical picture, actions, or exercises of agency (which are by nature active), are characterised as occurrent events. Boyle rejects certain aspects of this standard assumption, but accepts others. Specifically, his account is seemingly committed to the following key ideas:

- 1) An *event* (a dated happening taking place at a time or over time) is necessarily *occurrent*. This is in line with the standardly accepted view.
- 2) A *state* (which cannot be said to be presently occurring, or taking place) is necessarily *non-occurrent*. Again, this is in line with the standardly accepted view.
- 3) Something which is *active* may be either a *state* or an *event*. Given 1 and 2, it follows that something active can be either occurrent or non-occurrent. This is a rejection of the standardly accepted idea that all things which are active are necessarily events (and thus occurrent).<sup>29</sup>

What Boyle attempts to carve out is thus a distinctive a metaphysical category of *active, non-occurrent state*:

My claim is not that to believe something is to be occurrently up to something; it is that being occurrently up to something is not the only species of the genus: act, exercise of agency. (2009a: 137)

---

<sup>28</sup> One might draw a distinction between 'process' and state, rather than event and state. However, I take processes to be just one kind of event, i.e. one that unfolds over time.

<sup>29</sup> For our purposes we do not need to be concerned with Boyle's view on the following: whether occurrent things are necessarily events, whether non-occurrent things are necessarily states, whether passive things can ever be events.

On Boyle's view, a belief is conceived as belonging to the category of active state.<sup>30</sup> It is a state in that it is regarded as *a condition one is in* and not as something one *does*. But it is conceived as active in the sense that it is something one has a particular kind of 'intrinsic control' over (2009a: 137). Intrinsic control, on this understanding, consists in the capacity to dictate the existence of something and/or its nature. In this way one has intrinsic control over things that are exercises of one's agency. This is opposed to extrinsic control, which consists only in the capacity to influence or externally effect something. If one has extrinsic control over something, one influences it indirectly by doing other things to effect it (2009a: 125-128). So, according to this picture, I have intrinsic control over an intentional action such as waving my arm, but only extrinsic control over whether my friend waves their arm (I might prompt them to do so by waving at them, or instructing them accordingly). In this way, the control one has over one's own beliefs is taken on the reflectivist account to be analogous with the kind of control one has over one's intentional actions. I now attempt to clarify precisely what this means.<sup>31</sup>

According to Boyle, we can see what makes belief something we have intrinsic control over, and thus agential, by recognising a particular way in which it is similar to ordinary intentional action. Specifically, by recognising a way in which our relation to our own beliefs is structurally similar to our relation to our own intentional actions. This is initially motivated by appeal to the observation that I can reasonably be called upon to provide justifying reasons for both my intentional actions, and my beliefs. (2009a: 138). This fact is taken to demonstrate that we stand in a particular kind of relation to what Boyle calls 'the actuality' (i.e. existence) of the belief or the occurrence of the action. What follows is a reconstruction of this view as I understand it (drawing on Boyle: 2009a and 2011).

On Boyle's view, to presently be in a state of believing that *p*, one must regard *p* as 'to be believed'. In other words, one must *endorse* believing that *p*. One endorses believing that *p*, or regards *p* as to be believed, simply by regarding *p* as true: to affirm *p* as true is to be in a position of endorsing believing that *p*. Since regarding *p* as true just is what it means to believe that *p*, it follows that it is constitutive of belief that *p* that one endorses believing that *p*. More specifically, one's endorsing belief that *p* (in regarding *p* as true) *constitutes* one's belief that *p* (2009a: 140-1).

Ordinarily, one will regard *p* as true, and thus endorse believing that *p*, because one possesses reasons in favour of *p*. However, it is possible to have beliefs for which one does not possess

---

<sup>30</sup> Korsgaard (2009) advances a similar conception of belief.

<sup>31</sup> It is worth pointing out explicitly that Boyle's view relies on the idea that there can be non-intentional agential events, since belief is regarded as non-intentional. This is something I discuss in more detail in the next part of the thesis.

reasons. The crucial thing is that one cannot believe that  $p$  without also assessing  $p$  as true. In this way, belief that  $p$  is constitutively grounded by one's *ongoing* present assessment of  $p$  as true. In this way, the existence of the belief that  $p$  is dependent on an agent's continuing to endorse believing that  $p$ . Were one to cease regarding  $p$  as true, one would cease to endorse believing that  $p$ , and thus cease to believe that  $p$ .

How does the above show belief to be similar to intentional action? Just as believing that  $p$  requires regarding  $p$  as 'to be believed', performing an intentional action supposedly requires regarding that action as 'to be done'. When one is presently performing an intentional action ( $\phi$ ), it must be the case that one endorses doing so. If one does not endorse  $\phi$ ing, it cannot be the case that one is presently  $\phi$ ing. If I am intentionally waving my arm, it is because I regard waving my arm as to be done. If I do not endorse waving my arm, I cannot presently be waving my arm (at least not intentionally).

So, according to the view, both *intentional action* and *belief* require one to endorse *doing* the relevant action or *believing* the relevant proposition, respectively. In the case of belief, one endorses believing that  $p$  just in case one assesses  $p$  as true. In this way, the existence of both beliefs and intentional actions depends upon the subject's endorsement. In Boyle's terms, both beliefs and intentional actions are 'actualised' by this endorsement. It is for this reason that one can reasonably be called upon to answer the question of *why* one is  $\phi$ ing, or why one believes that  $p$ . Just as one is in a position of bringing it about that one is  $\phi$ ing by endorsing  $\phi$ ing, one is in a position of bringing it about that one believes that  $p$  by endorsing believing that  $p$ . In this way, one has *intrinsic control* over one's own beliefs, and one's relation to one's own beliefs is structurally analogous to one's relation to one's intentional actions.

How exactly does the supposed structural similarity between belief and intentional action demonstrate that belief is agential? Boyle explains it as follows.

It is this general relationship between endorsement and actuality, I suggest, that is the crux of rational agency, generically understood: where my present endorsement of X-ing is the ground of my present X-ing, in virtue of a capacity I possess to be through the former the source of the latter, there I am the agent of my X-ing, and X-ing is my act.  
(2009a: 141)

The thought thus seems to be that since rational agency for Boyle consists in presently providing the grounds for one's own believing or  $\phi$ ing, and since belief that  $p$  requires one to play this role by endorsing believing that  $p$ , then belief must be conceived as agential. The relationship between

endorsement and actuality is thus also conceived as the source of one's responsibility for one's own beliefs.

As should be evident from my reconstruction, Boyle's account of active belief rests on controversial ideas about what makes something an act. But I think that the idea can be made more intuitive by thinking about it along the following lines: in order for me to believe that  $p$ , it has to be the case that I regard  $p$  as true. Thus it seems that in order for me to sustain the belief that  $p$ , I must be, in some sense, in a sustained condition of regarding  $p$  as true. If we are compelled by the idea that a rational agent's assessment of some subject matter is something she is responsible for the way she is responsible for her intentional actions (and the fact that we can demand that a subject provide justification for her beliefs does seemingly support this), then it is, at least on some level, attractive to think that believing might be an exercise of agency, insofar as it is constituted by an ongoing assessment of  $p$  as true.

### 2.1.2 Self-knowledge as Constitutive of Active Belief

In the previous sub-section I attempted to clarify Boyle's view of belief as an exercise of agency. That part of the account conceived belief as a kind of non-occurrent mental act. In the present sub-section I discuss how this informs Boyle's conception of doxastic self-knowledge.

According to the reflectivist view, knowing that *I believe that  $p$*  is constitutive of believing that  $p$ . This is supposedly entailed by the fact that belief is an active state in the manner previously outlined: insofar as belief that  $p$  is constituted by one's ongoing endorsement of believing that  $p$ , belief that  $p$  is constituted by a *knowing* assessment of  $p$  as true. In this way, belief that  $p$  and knowledge that *I believe that  $p$*  are regarded as essentially *the same mental state*. To believe that  $p$  is to knowingly believe that  $p$  (2011: 228). In this way, reflectivism rejects the following standard assumptions: that believing that  $p$  and knowing that *you believe that  $p$*  are distinct states, and that reflection is required in order to acquire knowledge of one's own beliefs. Rather, belief that  $p$  and knowledge of that belief constitute a single state which is known to the subject without the need for reflection.

The idea that an act can be constituted by the agent's knowledge of it is reminiscent of Anscombe's account of practical knowledge (1963). In particular, it appears to align with her idea that knowledge of actions *formally describes* those actions (1963: 87). I accept something like Setiya's reading of this claim, according to which an act is formally described by the subject's knowledge of that act when it is constituted by her knowledge (2016: 158-9). Thus, knowledge of one's act would constitute the act only in certain cases, such as when one is 'paying, hiring, or marrying' (2016: 159). Boyle's view suggests that belief is also the kind of thing which is

constituted by a subject's knowledge of it, in that it is constituted by the subject knowingly endorsing it.<sup>32</sup>

A further important feature of Boyle's view is that, prior to *reflection* upon one's knowing belief state, doxastic self-knowledge is merely 'tacit'. By this it is meant that the belief is not yet explicitly self-attributed, in that the subject is not 'consciously thinking' (2011: 227) 'I believe that *p*'. Wherever one believes that *p*, one *implicitly* knows that one believes that *p*, and making a self-attribution is a matter of reflectively 'coming to explicit acknowledgment of a condition of which one is already tacitly aware' (2011: 227). Thus, the step from believing that *p* to judging that one believes that *p* involves making one's implicit doxastic self-knowledge explicit, i.e. conscious.

In order to explicitly self-attribute the belief that *p* to oneself, reflectivism holds that one need only reflect in the right way upon one's present knowing assessment of the subject matter *p*, since this constitutes one's belief, which itself constitutes knowledge of one's belief that *p*. Reflection in this context is understood in terms of *attention* (2011: 227-8). To make knowledge of her belief that *p* explicit, a subject turns her attention from the subject matter of her belief to her own engagement with that subject matter. Or, as Boyle puts it in a footnote (2011: 227, n. 4), she turns her attention 'from what [she is] representing to the mode of [her] own activity in representing it'.

On this view, knowledge of beliefs can be made conscious in ordinary cases through reflection upon them, and in cases where one is prevented from consciously knowing a belief, according to Boyle, both the belief and knowledge of that belief are present but 'not consciously accessible' (2011: 229). This has the surprising consequence that in cases of Freudian repressed beliefs and the like, it is not the case that one does not know what they believe, but rather that this knowledge is prevented from being made conscious by extenuating psychological factors. The knowledge is present within the subject, it is merely inaccessible.

So, the reflectivist view can be summarised as follows: doxastic self-knowledge is transparent in the sense that a subject need not reflect upon her own mental states in order to acquire knowledge of what she believes. This is because belief that *p* is metaphysically constituted by the

---

<sup>32</sup> For the sake of clarity (as well as brevity) this description of Anscombe's view leaves out the issue of intentionality. It can be noted however that for Anscombe, performing an action intentionally does not involve having some additional mental state accompanying that action which we could call an 'intention'. Rather, her thought seems to be that wherever one performs an action, insofar as it is 'the execution of an intention' (1963: 87), the action is intentional. So, an act like paying is intentional in that it *constitutes* an executed intention to pay. As Setiya puts it, 'it follows from the fact that one is doing A that one is doing A intentionally' (2016: 158). Since  $\phi$ ing intentionally involves knowing that one is  $\phi$ ing, it follows that in cases where one's  $\phi$ ing is constituted by an executed intention, one's  $\phi$ ing is constituted by knowing that one is  $\phi$ ing.



subject's knowing endorsement of believing that  $p$  – i.e. a subject must knowingly endorse believing that  $p$  in order to have the belief that  $p$ . According to Boyle's view of rational agency, on which something is an act insofar as its existence is dependent on one's endorsement, this makes belief both a standing state and a kind of non-occurrent act.

## 2.2 Objections

### 2.2.1 A Presentational Worry

The first objection I raise against Boyle's account is merely presentational. However, I believe that once observed this presentational issue serves to undermine some of the view's *prima facie* attractiveness. Through discussion of this initial problem I hope to clarify what claims the account is actually committed to, thereby opening the door to more substantive criticisms.

The issue is that Boyle's account at times uses language ordinarily employed to describe *occurrent acts* to describe the nature of *non-occurrent beliefs*. In doing so, the account suggests a more standard conception of agency than it in fact advances. Consider for instance the following passages:

...we have privileged self-knowledge of these sorts of attitudes because they are expressions of our capacity for rational self-determination. Their existence normally just consists in our knowingly assessing a certain matter in a certain way... their existence is normally constituted by our knowing assessment of the matter in question. (2011: 237)

[A reflectivist account] would explain my capacity for transparent knowledge of such states in terms of my ability reflectively to recognize an evaluation I already knowingly make, an evaluation my making which normally just constitutes my being in the relevant state. (2011: 237)

The discussion in these passages is intuitive and suggestive, but also somewhat misleading. It is claimed that belief consists in *assessing a subject matter*, that it is an *evaluation one makes*. This is the kind of talk we would usually employ to describe acts. To bring out the issue more clearly, we can observe that Boyle uses the perfective aspect in describing belief, presenting them as something that can be done. As we saw previously in Setiya (2008: 38), states are not the kind of thing that can be perfectly instantiated – they cannot be completed. Yet Boyle's descriptions slide between presenting belief as a state one is in and something one is engaged in doing.

Boyle's conception of belief is further muddled by the following kinds of analogy:

Consider a person who asserts P by saying, in an assertive tone of voice, 'P!' [...] there is a sense, palpable if difficult to explicate, in which he will normally know what he is up to[...] Furthermore, if he has mastered the needed vocabulary, he will on reflection be able to repeat his act in a way that makes its nature explicit — for instance, by saying 'I am claiming: P!' [...] The right account, surely, is that he all along had some sort of tacit awareness of what he was up to, and that this awareness, together with his mastery of certain classificatory expressions, allowed him to make his activity explicit, for us and for himself. All he had to do was to turn his attention to what he was already knowingly doing—i.e. to reflect. (2011: 227-8)

Again, it is easy to forget the details of Boyle's view when belief is compared to a 'doing'. Recall that the account explicitly rejects that belief is the sort of act in which one is 'occurrently up to something' (2009a: 137). Yet the above passage appeals to the idea that belief is similar to a case in which a subject 'all along had some sort of tacit awareness of what he was up to'. Granted, Boyle is discussing a very non-standard kind of act, and for this reason I think we can be forgiving if some of the expressions employed lend themselves more to a traditional conception of action – there is seemingly no natural way to speak of an act that is not a doing, a fact which highlights just how radical Boyle's account of belief is.

Nonetheless, it is important not to let the expression of the view obscure the specific account being proposed, according to which believing is not something a subject is presently doing. Rather, belief on this view is still understood as some way the subject's mind is – it is a state, or condition, she is in. In this sense at least, believing on Boyle's account is more similar to say, having green eyes, than to asserting p. On the reflectivist account, the crucial difference between a passive state and an active one consists in the particular 'intrinsic' control a subject has over active states, and not in the fact that one is something you do and the other is a state you are in.

Given the above considerations, I believe a more accurate way to express Boyle's conception of belief is as an act one *has* or *is in*, placing it in opposition to an act one does or performs. Presented like this the view looks altogether more daring, and indeed peculiar. In the next subsection I consider some more substantive problems with this metaphysical account.

### 2.2.2 The Analogy with Intentional Action

In this section I argue that one serious problem for the reflectivist view is that it seems to give a gratuitously complex explanation of the identified similarities between belief and intentional action. As a reminder, Boyle claims that belief and intentional action are similar in that a subject can be reasonably called upon to provide not merely explanatory, but *justifying* reasons for both.

So, just as one can be called upon to justify running down the corridor, one could likewise be called upon to justify believing that *there is an intruder in the building*. The clear implication here is that the subject has a similar kind of responsibility for both actions like running down the hall and mental states like believing that *there is an intruder*.

On some level Boyle is right to identify a similarity between belief and intentional action. It is true that we expect rational agents to be able to justify their beliefs by citing reasons in support of said beliefs.<sup>33</sup> In order for a rational belief to be sustainable over time and stand up to critical reflection, it must be the case that a subject can identify her reason(s) for having the belief. So, it seems that agents bear a particular kind of responsibility for their own beliefs, which suggests some capacity to exercise agential control over them. However, it does not necessarily follow from this that we must regard belief as *itself* an exercise of agency, as I will now attempt to demonstrate.

As we saw in Chapter 1, rational agents seemingly have a normative obligation to form beliefs on the basis of reasons which they regard as supporting these beliefs. And as Boyle observes, they also have an obligation to be able to report these reasons. If a subject cannot satisfy the latter obligation, that is, if she cannot provide a reason for her belief that *p*, then she fails to demonstrate that she has satisfied the former obligation.<sup>34</sup> The two obligations are connected in that the subject only has an obligation to provide justifying reasons for her belief *because* she has an obligation to believe on the basis of her agent's reasons. When a subject is asked why she believes that *p*, she is being prompted to demonstrate that she has met the obligation that she form beliefs on the basis of reasons. By fulfilling the requirement to cite her reasons for belief, she shows herself to have a rational, reason-responsive belief.<sup>35</sup> It thus follows that the obligation to have beliefs which conform to one's reasons is more fundamental than the obligation to provide reasons for one's beliefs, in that the latter obligation is dependent on the former.

The picture described above suggests that rational agents have a duty to report reasons for their beliefs because they are responsible in some sense for ensuring their beliefs conform to their

---

<sup>33</sup> It is also true that if a subject cannot provide justification for her belief, we expect that she will cease to have it.

<sup>34</sup> There is of course a difference between believing for a reason and being able to report one's reason for belief. One might well be motivated to believe that *p* for the reason *q* yet lack the ability to express (verbally or otherwise) what this reason is. However I assume that where a subject cannot identify, or cannot express, her own reason for belief, this ordinarily presents a problem for the rationality of the belief. Since the subject lacks evidence that her belief is well grounded, intuitively it seems she should no longer have it. In other words, I take it that there is some internalist condition on rational belief (see chapter on inferentialism for further discussion).

<sup>35</sup> Rational in one sense. As discussed in Chapter 1 the belief may still be globally irrational if the reasons motivating it are not normative reasons in favour of it.

reasons. This in turn suggests they have some form of agential control over what they believe. As a comparison, consider the following: you could not be obligated to ensure it won't rain this weekend, precisely because you have no control over the weather. Likewise, I have no obligation to ensure a lack of gender bias in the judging of this year's Booker prize, because I have no control over the judging process. In the same way, a subject could not be obligated to ensure her beliefs conformed to her reasons if it were not the case that she had some control over this. However, it does not follow from the fact that she has some form of agential control over what she believes, and can thus be called upon to justify her beliefs by providing reasons, that belief must itself be understood as an act.

To see why Boyle's observation doesn't require us to think of belief as an act, we can consider the following kinds of cases: just as a subject can be called upon to provide justifying (as opposed to merely explanatory) reasons for running down the hallway, or for believing *there is an intruder in the building*, a subject can be called upon to provide justifying reasons for things like having her bedroom window open. Yet it is not the case that having one's bedroom window open is an act. The subject's condition or state of having her window open is only a product or consequence of the subject's exercising agency, and not itself an exercise of agency. The reason she can be called upon to justify such things is that she engaged in action to bring them about, or else she allowed them to come about by deciding not to take action to change them. Expressed using Boyle's earlier terminology, these are things over which the subject exercises *extrinsic* rather than *intrinsic* control. They are brought about by the subject doing things but they are not themselves things the subject does. Yet she can still be held accountable, and thus requested to justify, things for which she is only extrinsically responsible.

One could claim, in response to the above, that what the subject in the above cases is actually being asked to justify is not so much a state of affairs, or a condition she is in, but the action(s) which brought these about. In this case, justification is being requested for something the agent is intrinsically responsible for. However, this observation only serves to harm Boyle's comparison. If we can think of the request to provide justifying reasons for having an open window as a request for justification for performing the *act* of opening the window, then we can think of the request for justifying reasons for believing that *p* as a request for justification for *forming the belief that p*. Yet Boyle does not want us to think of our responsibility for belief as stemming from the event or process of belief formation. Rather on his account responsibility for belief comes from belief itself being an exercise of agency.

On the assumption that Boyle would not want to treat the kinds of cases I have mentioned as non-occurrent acts, then the similarity he identifies between intentional action and belief cannot

be treated as good evidence for regarding belief as a kind of mental act. In other words, the similarity need not be explained in terms of intrinsic control. Of course, those cases could not qualify as agential on Boyle's account since they are not grounded in the appropriate way – i.e. they are not *dependent* on the subject's endorsement. But this account of the agentialness of beliefs is proposed partly as an explanation of the apparent symmetry with intentional action, and insofar as the same symmetry exists with something that is not dependent on the subject's endorsement, the account is unnecessary.

So, if our doxastic obligations don't need to be explained in terms of intrinsic control, how else might we do justice to them? It seems true after all that we have an obligation to justify even those beliefs which we formed unconsciously, or without mental action. One way of explaining our responsibility for our own beliefs is by observing that they are importantly connected to our capacity to engage in various forms of mental action, even if they are not themselves acts. In other words, we can understand our responsibility for them as stemming from a kind of extrinsic control, comparable to that which grounds our responsibility for having a window open, etc. For instance, one way in which they are connected to agency is in that they can be brought about through conscious rational reflection on a subject matter. Another is in the fact that they can be evaluated and revised through critical reflection. Ultimately, I think it is unlikely that a subject has the same kind of responsibility for all of her beliefs, precisely because her accountability for her various commitments can be grounded by various kinds of mental action and agential capacities.

However, Boyle attempts to show that a subject's responsibility for her own beliefs must be 'characteristically agential' (2009: 121) if we are to avoid thinking of the subject as alienated in some way from her belief. That is, he tries to argue that we can only make sense of the kind of accountability we have for our own beliefs by understanding it as grounded in action. To demonstrate this point he suggests that responsibility for belief, if it was grounded in the fact that belief is *produced* by action, would have the result that a subject is only *indirectly* responsible for her beliefs, in a manner similar to how she is responsible for something she owns (in this case a bicycle):

I am not responsible for my beliefs merely in the way I am responsible for my bicycle: I am not just responsible for having left them in a certain location, as it were. That gets the temporal aspect of my responsibility for them wrong in a fundamental way. I am not directly accountable only for *having arrived* at certain beliefs, and hence indirectly accountable for the situation my beliefs are in. I am directly accountable for *believing* what I do, no matter what the antecedents of my psychological condition may have been. I am not merely expected to be able to say why *I came to* believe something, but

why I believe it [...] this suggests that our discretion over our beliefs consists, not merely in our ability to control their installation and removal, but in something about the nature of our believing itself. (2009: 136)

So, Boyle emphasises the idea that what we are called upon to explain, when asked to justify a belief, is not why we formed the belief, but why we presently have the belief. On his understanding, thinking of responsibility for belief as deriving from the fact that beliefs are produced by mental actions cannot do justice to this. For Boyle, to deny that responsibility for beliefs derives from the agentialness of belief leaves us in the position of regarding beliefs as things for which we are only indirectly responsible, as we are responsible for something like a bike. This kind of responsibility, he argues, leaves us accountable merely for forming and removing our beliefs.

However, the analogy is misleading. It is *not* the case that denying the notion of belief as action requires committing to a picture on which one's responsibility for belief is essentially similar to one's responsibility for a bike. There is a clear sense in which my capacity to perform occurrent mental actions, such as conscious judgments, means that I have a more direct kind of control over my beliefs than I have over my bike. My beliefs are directly, rationally tethered to my occurrent mental actions, including my conscious judgments.<sup>36</sup>

When I make a judgment, this event will normally be token identical with a change in my mental states.<sup>37</sup> To occurrently affirm *p* will, under normal circumstances, not merely cause me to acquire the belief that *p*, as moving by bike will cause it to be in a new place, rather that mental event will itself *constitute* an event of belief formation. In this way there is no gap between the act and the thing brought about by that act. Moreover, I am normally disposed to immediately revise beliefs in light of changes in my reasons, such as new information. In this way, standing commitments are constantly sensitive to, and apt to be shaped by, my agent's reasons.

How do my claims here serve as a response to Boyle? Boyle points out that when asked to give reasons for a belief, 'I am not merely expected to be able to say why I came to believe something, but why I believe it'. I think this is broadly right. We are usually being asked to give reasons for presently holding the belief.<sup>38</sup> However, even a belief conceived as a *non-agential* state is, at any given moment, sensitive to one's reasons, insofar as it is sensitive to

---

<sup>36</sup> I will say more about the nature of judgment in my positive chapters. For now I offer a brief sketch of the kind of mental act I take judgments to be.

<sup>37</sup> Jenkins, 2018.

<sup>38</sup> Though presumably it will sometimes be appropriate to give a story about how you reached a judgment.

one's active assessments. The fact that I presently have the belief that  $p$  is itself an expression of my responding to reasons in a particular way. If at some point I came to think my reasons no longer supported  $p$ , I would, all things being equal, cease to believe that  $p$ .

Thus when asked why I believe that  $p$ , I am being invited to reflect upon and report the reasons to which my belief presently conforms. And of course, if I find myself to have a belief which is not rationally tethered to my reasons, then I have a rational obligation to revise it. This is all to say that we need not think of belief as itself an act in order to think of it as something for which a subject is presently accountable (in that she is accountable for the belief itself and not merely for forming it). What Boyle points out suggests that we expect beliefs to be constantly sensitive to our assessments of our reasons, but such sensitivity does not entail activity.<sup>39</sup>

Boyle's claim that we must understand beliefs as actions in order to make sense of our responsibility for them thus appears unfounded. Even if we deny that beliefs are actions, it cannot be the case that we are responsible for them in a way that is analogous to owning a bicycle. This is due to the especially direct rational connection between mental action, i.e. occurrent reasons assessment, and standing beliefs. Moreover we can retain the idea that we are accountable for our beliefs such that we can be called upon to justify having them, as opposed to simply arriving at them, without regarding them as exercises of agency. In part two of this thesis I will develop a positive account of the responsibility we have for beliefs in greater detail.

### 2.2.3 The Over-Reflective Nature of Belief on the Reflectivist View

As I demonstrated in the first section of this chapter, the reflectivist account is dependent on Boyle's unusual metaphysical conception of belief. It is only because beliefs are understood as acts on this view that they can be conceived as necessarily known to the subject who has them. In this subsection I consider a problem for Boyle's metaphysical conception of belief which has serious ramifications for his epistemology. Specifically, I advance an objection to Boyle's notion that belief that  $p$  is constituted by one's knowing endorsement of believing that  $p$ . I begin

---

<sup>39</sup> Of course, not all of our beliefs are formed through mental action. I take there to be a question over the nature of our responsibility for our unconsciously acquired or purely dispositional beliefs. In this thesis my focus will be on the responsibility we have for beliefs formed through judging, since I take transparent warrant for belief to derive from the nature of conscious judgment. It is worth noting however that even beliefs which are not acquired through judgment are importantly connected to mental action, in that believing that  $p$  ordinarily disposes one to engage in particular mental actions, such as judging that  $p$  (McHugh, 2011: 264-265)

however with a reminder of two key ways in which the reflectivist account of belief comes into tension with the standardly accepted thinking about belief and knowledge of beliefs.

As I previously observed, the following are standard assumptions, in both philosophy and non-academic thinking: first, believing something and knowing you believe something are two distinct states, and second, in order to gain awareness of our own beliefs, we must engage in some kind of reflective process.<sup>40</sup> Reflectivism rejects both of these standard ideas. Instead, it takes believing that  $p$  and knowing that one believes that  $p$  to be a single mental state, and as a result denies that reflection of any kind is required to gain knowledge of one's own beliefs. The reflective process described by Boyle's view refers to the way in which we *consciously manifest* our doxastic self-knowledge, and *not* how we acquire it.

We should be cautious to keep in mind that Boyle wants to claim beliefs are tacitly known to subjects in virtue of the peculiar metaphysical nature of belief, and not because they are produced by engaging in less controversial forms of mental action, as we might find in conscious deliberation. It needn't be that one consciously makes an assessment about a subject matter for her to have a belief that is known to her. The idea is that all beliefs, regardless of whether their *formation* involved exercises of mental agency, are themselves the kind of things which are agential, and thus known to their subject. Reflectivism is therefore committed to the following, somewhat dubious, notion: in cases where a subject has a belief, but either struggles to self-attribute any belief, or falsely self-attributes a different belief from the one she in fact possesses, the subject *has* knowledge of the correct belief, she is simply unable to make this knowledge explicit (2011: 229).

A further seemingly unintuitive effect of construing all beliefs as tacitly, or implicitly, known by their subjects is that we are taken to have awareness not only of beliefs which we fail to identify, but also of the innumerable implicit, dispositional beliefs manifest by our actions and behaviour. A subject might have the following sorts of beliefs: that her pen will fall to the floor if she drops it, that there are no orangutans in her back garden, and that the ground she is standing on is solid. And she may very well be disposed to explicitly affirm the propositional contents of these beliefs. Yet while it is difficult to imagine that she could know herself to have all of these beliefs, this is precisely what the reflectivist account invites us to accept.

---

<sup>40</sup> On traditional philosophical accounts of self-knowledge this reflection consists in a kind of observational process. On most transparency accounts it is a matter of reflecting upon the subject matter of one's own beliefs.



The reflectivist account is committed to these controversial ideas because of the manner in which it construes belief as active. Recall that the agentialness of belief is taken to consist in the fact that belief that  $p$  is constituted by one's 'endorsing' belief that  $p$  (i.e. regarding  $p$  as 'to be believed'), which itself is constituted by regarding  $p$  as true. I accept that belief is constituted by being in a sustained state of regarding  $p$  as true. What I take issue with is the idea that for a rational agent to regard  $p$  as true *constitutes* what Boyle calls endorsing belief that  $p$ .

As I understand it, to endorse a belief on the reflectivist view is to have an attitude, constitutive of belief that  $p$ , with the content  *$p$  is to be believed* (or ought to be believed). I will assume that this attitude which constitutes belief is itself a belief, meaning that one cannot believe  $p$  without also believing that  *$p$  ought to be believed* (which would constitute belief that  $p$ ). This would need to be the case even in instances of non-deliberatively (or non-consciously) acquired dispositional belief. So, a subject who had the unconscious belief that *the floor is solid* would, in virtue of having this belief, also have the belief that *it ought to be believed that the floor is solid*.

The problem with this notion of endorsement is that it seems to make belief overly complex and self-aware in reflective beings. It supposes that having an affirmative attitude towards a subject matter is identical with, insofar as it is constituted by, having the attitude that *one ought to have an affirmative attitude towards  $p$* . In this way, what is seemingly a fairly complex second-order state apparently collapses into the first order state of believing that  $p$ . Yet it seems unlikely that our non-reflective dispositional beliefs are constituted by this kind of conceptually sophisticated thinking. Certainly it does not appear essential to belief that one regard the belief as appropriate or required. Part of Boyle's motivation for defending such a sophisticated account of belief is that he views it as necessary to explain a specific symmetry between intentional action and belief, i.e. that a subject can be called upon to provide justifying reasons for her belief. But as I discussed in 2.2., more straightforward explanations are available.

It is unlikely that the reflectivist would want to accept the version of the view I have assumed, on which endorsement is understood as itself a belief. This is because, aside from being demanding, such an account would seemingly result in a vicious infinite regress, due to the constitution relation between endorsement and belief. However, if endorsement consists in something other than belief, it is unclear what this could be.

In the next section I consider further problems related to Boyle's demanding account of belief. These arise from the possibility of certain kinds of epistemic failure, specifically self-ignorance, and doxastic akrasia.

#### 2.2.4 The Problem of Epistemic Failure

As previously noted, one unintuitive feature of the reflectivist account is that it suggests that when a subject has a belief which she is unable to identify, for instance in the case of repressed beliefs, the subject *has* knowledge of this mental state, she is simply unable to access it. That is, she is unable to make this knowledge *explicit*. A serious problem with this is that it looks to be an *ad-hoc* claim. That is, it seems to be a claim made not because it is generally plausible, or because there are independently good reasons to make it, but because it is necessary to preserve the wider account.

In this section I consider the issue in more detail and demonstrate that there are reasons to be worried about this aspect of the reflectivist account beyond its general lack of intuitive appeal. I then consider whether the possibility of a different kind of epistemic failure, akratic belief, presents a problem for the view. I identify doxastic akrasia as the phenomenon of having beliefs which one would not consciously avow or endorse.

##### 2.2.4.1 Self-Ignorance in the Case of Repressed Belief

In particular, tacit self-knowledge of the kind reflectivism describes lacks a feature ordinarily distinctive of tacit belief: it does not serve to explain the behaviour of the subject.<sup>41</sup>

What does it mean to say that tacit belief ‘explains’ behaviour? I take it that in standard cases, where a subject has a tacit belief (i.e. a belief which is not conscious) this belief will be manifest in the subject’s behaviour. In this way, tacit knowledge plays an *explanatory* role, insofar as it explains why the subject acts in the way she does. For example, a subject’s belief that *it will snow today* might be manifest in her putting on snow boots. Or, her belief that *the door she is about to push open is heavy* might manifest in her raising and bracing her arms as she approaches said door. Moreover, were she *not* to behave in ways that manifest such beliefs, e.g. if she put on loafers rather than snow boots, or if she did not brace before going to open the door, then we would have reason to doubt whether she in fact believed that *it was going to snow*, or that *the door is heavy*.

The reflectivist account suggests that wherever a subject has the belief that *p*, she also has the tacit belief that *she believes that p* (which is knowledgeable). Now, if this is the case, then we

---

<sup>41</sup> I will utilise the notion of tacit knowledge in my own proposed account, where it will be discussed in greater detail (see REF). In this later discussion I explain that there are *three* standard features of tacit knowledge.

should expect that the subject's second order belief<sup>42</sup> will be manifest in her behaviour. That is, we would expect it to somehow *explain* her behaviour. Yet it is unclear what behaviour the reflectivist could point to, particularly in the case of repressed beliefs, as manifesting tacit knowledge of one's own beliefs. While repressed beliefs may *themselves* be manifest in behaviour, e.g. someone who has the repressed belief that they are going to become sick might start adopting extreme lifestyle changes to improve her health, there is usually nothing to suggest that these subjects *know* that they have the repressed belief.

In fact, even if we grant that there are some cases of tacit belief where the relevant epistemic commitment is not manifest by the subject's behaviour, the subject in possession of a *repressed* belief acts in a way that seemingly manifests a *lack of awareness* of her own belief. Ordinarily, the way we manifest our doxastic self-knowledge is by correctly identifying and reporting it. The fact that the subject with a repressed belief *cannot* report having the belief seems to demonstrate that she lacks knowledge of it. To put this point in slightly different terms, the behaviour of the subject in these cases is better explained by her *lacking* self-knowledge than it is by her possessing it.

#### 2.2.4.2 Akratic Belief

A related objection to the reflectivist conception of tacit doxastic self-knowledge arises from the possibility of akratic belief. I briefly touched on akrasia in 1.1. There, it was claimed that beliefs which do not conform to one's reasons seem to be irrational in a way similar to akratic action. Just as one who acts akratically, arguably,  $\phi$ s without taking herself to have reason to  $\phi$ , the subject who *believes* irrationally (in the agentalist sense) has a belief which does not conform to her agent's reasons (those reasons she identifies as normative).

Here, I am interested in beliefs that are irrational not just in that they do not conform to the subject's reasons, but in the sense that the subject herself consciously rejects their contents. An akratic belief, according to the present conception, is one which the subject possesses while simultaneously believing it to be an attitude she ought not to have. Such beliefs are ordinarily beliefs which a subject would be unlikely to consciously avow, yet which are nonetheless recalcitrant.

Here are two examples of the kind of beliefs I have in mind:

---

<sup>42</sup> Second order in the sense that it is a belief about her belief. On Boyle's account at least it would not be second order in the sense that it is a separate mental state.

Roland is about to have surgery, when he discovers that the surgeon performing his operation is a woman. Roland becomes nervous, as he has the belief that *women are less competent at performing complex medical procedures than men*. But Roland is also a feminist, and believes this belief is one he ought not to have. Indeed, when prompted to reflect on the matter he would reject the proposition that *women are less competent surgeons* as false. Nonetheless, his belief is the deeply ingrained product of being raised in a patriarchal society, and as such is particularly recalcitrant.

Maude is turning 30, and her friends have thrown her a party. Maude feels uncomfortable with this, as she believes that *she is undeserving of having her birthday celebrated*.

However, Maude is also suffering from depression. When she discusses her feelings about the party with her counsellor, Maude comes to believe that she ought not to believe that *she is undeserving of having her birthday celebrated*. Nonetheless, she fails to shake the depressive belief.

In both of these cases, as I have described them, the subjects have beliefs which they believe they ought not to have. The question I want to address now is whether the possibility of such cases is a problem for reflectivism. *Prima facie*, they do present a problem, since the reflectivist account holds that belief that *p* is constituted by one's *endorsing* having the belief that *p*, and yet here we have two cases in which the subjects have beliefs which they do not endorse.

One way in which the reflectivist could respond to the worry is by arguing that the fact that Roland and Maude do not approve of their own beliefs does not itself show that they fail to endorse them in the sense Boyle has in mind. For instance, it could be argued that Roland and Maude are both in some state of conflict brought about by their respective emotional biases, in which they regard *p* as something to be believed (thus fulfilling Boyle's conception of endorsement) *and* regard it as something they ought not to believe.

However, this picture seemingly raises a paradox. Recall that I have understood Boyle's conception of endorsement as believing that *p is to be believed*. Given this fact, accepting the response I have proposed would require one to accept that in all cases of akratic belief, the subject not only has a belief of which they themselves do not approve, but they also simultaneously believe both that *p is to be believed* and that *p is not to be believed*.<sup>43</sup> Even if we accept that such a state of affairs is possible, on the reflectivist account it presents an extremely

---

<sup>43</sup> This is essentially the 'static' paradox levelled at accounts of self-deception on which the self-deceived subject is regarded as having both a true second order belief about *p* and a false first order belief (Mele, 2001).

complex picture of doxastic akrasia, according to which the subject seems guilty of two discrete epistemic failings.<sup>44</sup> These failures are as follows:

The subject has a belief that *p* of which she does not approve. This is the key epistemic failure of doxastic akrasia. The subject believes that *she ought not to believe that *p**, but believes that *p* anyway.

The subject endorses both believing that *p* and not believing that *p*. As a result of the reflectivist regarding endorsement as constitutive of belief, the subject would be required to knowingly endorse believing that *p* while also believing that *p ought not to be believed*.

The reflectivist view thus entails an improbably complex account of doxastic akrasia, according to which the subject experiences multiple epistemic failures every time she believes against her better judgment. Yet, as in the case of repressed belief, there seems little reason to accept that this is what happens in cases of doxastic akrasia, beyond making the phenomenon fit with the reflectivist account.

## 2.3 Conclusion

The reflectivist account develops key ideas from Moran's agentialist account. In doing so, it addresses the questions left open at the end of the last chapter: reflectivism provides explanations for the nature of our awareness of our own rational assessments, how this awareness warrants knowledgeable self-ascriptions, and how we go about making self-ascriptions.

On the reflectivist account, a subject knows her own beliefs because they are her *acts*. So, the account doesn't need to explain the connection between agential reasons assessment and belief – beliefs are themselves agential exercises. For this reason, the account claims that beliefs are known to their subject. Self-knowledge is thus not merely understood as non-inferentially grounded, it is actually taken to be constitutive of belief. One tacitly self-ascribes belief just in virtue of believing. *Explicit* self-ascriptions are made by reflecting, i.e. *attending* to one's condition of belief (which one is in knowingly) in the right way.

However, Boyle's own epistemology faces problems as a result of its metaphysical picture of belief. As we saw, reflectivism begins with the controversial idea that belief represents a form of non-occurrent action. I attempted to demonstrate that Boyle's argument for thinking of belief as

---

<sup>44</sup> Note that these failures follow even if one rejects that endorsement consists in belief that *p is to be believed*.

## Reflectivism

an act fails to convincingly show that our responsibility for beliefs can only be explained by conceiving of belief as action. I then argued that by construing belief as constituted by knowingly endorsing belief that *p*, reflectivism makes belief overly complex. This consequently results in an overly complex account of certain epistemic failures.

There is nonetheless something appealing about Boyle's notion that some of our mental episodes are tacitly known to us in virtue of being exercises of our agency. However, the account is less convincing in its assertion that beliefs constitute such exercises of agency. In Part 2 of the thesis I argue that occurrent, conscious judgments, rather than beliefs, are mental events which are tacitly known to us. As a result one is in a position to self-ascribe the belief that *p* whenever one judges that *p*.

For now, I turn my attention on one more transparency account. In the next chapter I consider the inferentialist account proposed by Alex Byrne (2011, 2018). This view presents an alternative to the accounts considered in Chapters 1 and 2. Both agentialism and reflectivism take our warrant for transparent doxastic self-knowledge to be explained by our capacity for rational agency. Specifically, by the fact that our beliefs are expressions of our capacity to make rational assessments of our reasons. On Moran's agentialism, beliefs are the *product* of active rational assessment. On reflectivism, beliefs are *themselves* active rational assessments. However, the inferentialist account rejects that rational agency plays any important role in our transparent self-ascriptions. Instead, the account suggests that transparent self-knowledge is simply the product of a reliable, safe inference from *p* to *I believe that p*.

## Chapter 3      Inferentialism

This chapter will consider Byrne's inferentialist transparency account of doxastic self-knowledge (2011, 2018). Inferentialism represents a significant departure from the views developed by Moran and Boyle. As we have seen, those views understand doxastic transparent self-knowledge as being made possible by the relationship between belief and rational agency. In contrast, inferentialism does not appeal to the role of rational agency in belief. Instead, Byrne's view construes transparent self-knowledge simply as the product of an inference from some fact about the world,  $p$ , to the conclusion *I believe that p*.

Byrne's account is the most fully developed existing view that does not rely on rational agency to explain the distinctive transparency of doxastic self-knowledge.<sup>45</sup> For this reason, considering whether inferentialism can satisfy the previously outlined desiderata for a successful account of self-knowledge will aid us in assessing the importance of rational agency. If Byrne's account can give a satisfying explanation of the phenomenon of transparency alongside a plausible explanation of how transparent self-knowledge is warranted, this would suggest that the views developed by Moran and Boyle overstate the role of rational agency in transparent self-knowledge.

However, I will claim that inferentialism falls short of giving a plausible explanation of the epistemic warrant for transparent self-knowledge. Thus, like agentialism and reflectivism, inferentialism fails to satisfy the second desideratum for a successful account of self-knowledge.

Sections 3.1 – 3.3 will be concerned with elucidating the view under consideration. In section 3.4 I consider objections to the view. These will be primarily concerned with developing the idea that the relevant inference is irrational, a criticism originally proposed by Boyle (2011). I will argue that, since the inference is not rational, it cannot be understood as conducive to knowledgeable self-ascriptions.

### 3.1      The Doxastic Schema and BEL

Byrne's inferentialism can be understood as addressing an ambiguity in Moran's transparency account. Recall the central claim of Moran's view: *insofar as an agent is rational, she is in a*

---

<sup>45</sup> At least not in any demanding sense. I take it that the relevant kind of inference would involve some minimal form of mental agency. The point is that Byrne's account doesn't appeal to the fact that beliefs are expressions of rational agency to explain the transparency of doxastic self-knowledge.

position to answer the question ‘what do I believe about  $p$ ?’ by answering the world-directed question ‘is  $p$  true?’. In this sense the first-person question is ‘transparent’ to the world-directed question. According to Moran, when a subject acquires doxastic self-knowledge transparently, she answers the former by engaging in *deliberation* over the latter. In this way it is non-voluntaristically ‘up to’ her what she believes (2001: 66-67). What Moran’s account does *not* provide is an explicit statement of how exactly a conclusion about  $p$  can ground the belief *that I believe that  $p$* , though Moran does state that this event is ‘immediate’ in the sense that it is non-inferentially grounded (2001: 90).<sup>46</sup>

However, Moran’s basic transparency claim, that a subject can learn what she believes about  $p$  by answering the question of whether  $p$ , does not depend on the grounding relation between world-directed judgment and self-attribution being non-inferential. As such it is possible to hold onto the key positive statement of Moran’s view without conceding that self-knowledge of beliefs is acquired non-inferentially, and this is essentially what the inferentialist account does. Specifically, it construes the transparency method as reasoning in line with the ‘doxastic schema’, a model of inference proposed by Gallois (1996). For a subject to reason according to the schema is for her to draw an inference from the premise that  $p$  to the conclusion *I believe that  $p$* . The model is presented in the following way:

$p$

---

I believe that  $p$

Byrne takes the inference presented by Gallois’s schema to be underwritten by an epistemic rule he labels BEL (2018: 102, 2005: 95).

BEL: if  $p$ , then believe that you believe that  $p$

This rule, rather than the schema itself, is the central focus of Byrne’s most recent discussion. Since inferring according to the schema is how one follows the rule on Byrne’s account, references to ‘following BEL’ in the present discussion should be treated as roughly equivalent to ‘reasoning in line with the doxastic schema’.

---

<sup>46</sup> I take this ambiguity to be a significant problem for most agency-based views. As we have seen Boyle’s reflectivist account also struggles to explain precisely what happens when a subject transitions from making a judgement about  $p$  to explicitly self-attributing the belief *that  $p$* .



While I will not discuss in any detail Gallois's own inferentialist transparency view, which is both complex and obscure, it is interesting to note a few things about his account. Gallois thinks that the doxastic schema is knowledge conducive, and that we must reason according to the doxastic schema in order to maintain a rational world-view. He attempts to show why it is necessary for rationality (1996:76), but doesn't actually try to demonstrate how it is knowledge conducive. The result of this is that his view leaves open the 'apocalyptic possibility', to use Byrne's phrasing, that rationality requires us to sacrifice self-knowledge (Byrne 2018: 78). This fact alone gives me adequate reason to dismiss Gallois's view for now, since like Byrne I am interested in finding an account that explains what makes transparently acquired beliefs warranted such that they count as knowledge, as opposed to merely rational.<sup>47</sup>

Byrne's account thus aims to demonstrate what Gallois's account fails to show – that the doxastic schema is knowledge conducive. The account also aims to show that taking self-knowledge to be acquired in this way has a number of compelling benefits (2011: 207). Crucially it is supposed to explain the 'peculiar' and 'privileged' nature of the access we have to our own states. In the next section I clarify this terminology.

### 3.2 Privileged and Peculiar Access

To think of self-knowledge as *privileged* in the inferentialist sense is to take it to be epistemically secure in a way that other kinds of knowledge are not (2011: 202, 2005:80). By this we mean that the beliefs an agent holds about her own states are generally more likely to be correct and therefore to count as knowledge, than beliefs about other agents' mental states. This need not entail that we are infallible or omniscient regarding our own mental states, but it does rest on the assumption that we are better positioned to 'get it right', as it were, when it comes to our own states. This is what Byrne refers to as the 'privileged access' we have to our own mental states.

'Peculiar access', on the other hand, refers to the fact that agents are seemingly able to acquire knowledge of their own mental states in a way that they cannot acquire knowledge of others' mental states (2005: 81, 2011: 202). More specifically, they can learn what they believe without appealing to the kinds of evidence they rely on to learn about others' beliefs.

According to Byrne, it is peculiar access which explains *why* we have privileged access.<sup>48</sup> Peculiar access concerns the special method through which we access our first order beliefs, while

---

<sup>47</sup> For a more detailed discussion of Gallois see Byrne (2018: 77-9).

<sup>48</sup> Though we should note that Byrne takes the features to be independent insofar as neither is necessarily entailed by the other (2005: 81, 2011: 202).

privileged access concerns the special epistemic security of our second order beliefs. I will now discuss how Byrne's transparency account purports to do justice to these features.

### 3.2.1 Peculiar Access

As stated, peculiar access refers to the fact that agents' have a way of acquiring knowledge about their own mental states that they cannot use to gain knowledge of others' mental states. The easiest way to see this is by thinking about what is *not* involved in acquiring self-knowledge.

If I want to learn, for example, whether I believe it will be a hot day, I needn't consider evidence like whether I'm wearing shorts, or whether I put sun cream on before leaving the house. But if I want to learn whether Charlie believes it will be a hot day, I rely on either asking him, or making such observations about his behaviour. The fact that I need not rely on this kind of evidence to learn of my own beliefs, while it is essential for me to learn about other people's beliefs, suggests that I have some form of distinctively first-personal access to my own doxastic attitudes.

Byrne's account purports to explain this special kind of access, since the method he describes does not involve inference from evidence about oneself<sup>49</sup>, and can only be used to learn of one's own beliefs (2011: 207, 2018: 108-9). According to Byrne, this is because the epistemic rule that underwrites the doxastic schema, BEL, is 'strongly self-verifying': to recognise *that p* is to believe *that p*, thus, since following BEL requires recognising *that p*, one can only follow BEL when one believes *that p*. In other words, reasoning from *p* entails belief in *p*.

If one reasons in accord with the doxastic schema [thereby following BEL], and infers that one believes that *p* from the premiss that *p*, then one's second order belief is true, because inference from a premiss entails belief in that premiss. (Byrne, 2011: 206)

So, following BEL guarantees that the second order-belief I form as a result will be true. It should be noted that this is asserted with a caveat, that it is theoretically possible for one to follow BEL and end up with a false second order belief in cases where one changes one's mind during reasoning (2018: 104). However, Byrne claims that since the chain of reasoning captured is so short, it is highly unlikely that this will happen, thus he operates for the most part as though the possibility does not exist. I will do likewise in my discussion, mainly for the reason that there are numerous more substantive and interesting issues facing the account, even if we ignore this qualification.

---

<sup>49</sup> Indeed there is a very real sense in which it does not involve inferring *from evidence* at all. This will be discussed later in the chapter.

### 3.2.2 Privileged Access

While peculiar access refers to the idea that we have a distinctively first-personal way of learning about our own beliefs, privileged access refers to the notion that our doxastic self-knowledge is more secure than other kinds of knowledge. That is, there is less chance of going wrong than when it comes to other people's mental states. To capture this, Byrne's account once again appeals to the idea that BEL, the epistemic rule underwriting the doxastic schema, is self-verifying. To reiterate, this is the idea that following BEL requires one to have the belief that  $p$ , and thus when one follows BEL one is guaranteed to arrive at a true second order belief about what one believes (2018: 104).

According to Byrne, the fact that BEL is self-verifying means that it is better at producing knowledge than rules whose consequents concerns other people's beliefs (2018: 109). We can try to see this contrast by thinking of a knowledge conducive rule that does concern another subject's beliefs<sup>50</sup>, and which one might plausibly follow:

SUN CREAM: If Charlie is applying sun cream, believe that he believes it is sunny.

SUN CREAM is not self-verifying, since it doesn't need to be the case that Charlie believes it is sunny for me to follow SUN CREAM. I can follow it and end up with a false belief about what Charlie believes. Because of this, such a rule produces beliefs that are less *safe* than those produced by BEL. Byrne characterises safety by approximating a formulation from Williamson (2000: 101-2): 'one's belief that  $p$  is *safe* just in case one's belief could not easily have been false' (2018: 110).

Byrne demonstrates that BEL produces safer beliefs than a rule like SUN CREAM by considering three types of error one might make that would lead to one having a false belief. I will first apply these to SUN CREAM. In each of the following cases,  $p$  refers to the fact that Charlie believes it is sunny. Conditions C refer to Charlie's applying sun cream.

Type I:  $\sim p$ , and one falsely believes that conditions C obtain, thereby believing that  $p$ .

I might see Charlie applying moisturiser and mistakenly think it is sun cream, leading me to believe falsely that he believes it is sunny.

Type II:  $\sim p$ , and one truly believes that conditions C obtain, thereby believing that  $p$ .

---

<sup>50</sup> Byrne in fact contrasts BEL with the following rule: DOORBELL: if the doorbell rings, believe there is someone at the door. But I feel it will be more helpful to consider a rule that refers to another subject.

Charlie may have mistaken his sun cream for moisturiser, and unknowingly protected himself from UV rays while merely intending to hydrate his skin. In this case I am again lead to a false belief in spite of the fact that Charlie really is applying sun cream.

Type III:  $\sim p$ , and one believes that  $p$ , but not because one knows or believes that conditions C obtain.

Perhaps I have absolutely no belief about whether Charlie is applying sunscreen, but have ingested a pharmaceutical that has the side effect of leading me to spontaneously acquire some beliefs at random, including the belief that *Charlie believes it is sunny*.

According to Byrne, there are realistic cases in which all of the above errors are nearby possibilities. In such cases, one's true belief that *Charlie believes it is sunny* would not, according to the safety condition, count as knowledge. However, when one follows BEL (or at least *tries* to follow BEL), Type I and II errors are supposedly impossible. If, for example, I follow BEL to come to believe that *I believe it is sunny* (thereby following this rule: if it is sunny, believe that I believe it is sunny), I cannot make a Type I error because I cannot falsely believe that *it is sunny* without it being true that *I believe that it is sunny*. And I seemingly cannot make a Type II error because I cannot truly believe that *it is sunny* without it also being true that *I believe that it is sunny*.

On Byrne's view, only Type III errors are a problem for BEL (2018: 111). Theoretically, there are circumstances under which I could come to believe that *I believe it is sunny* without actually believing that *it is sunny* (e.g. as a weird effect of some drug, or in Byrne's own example, too much coffee). Byrne claims that, assuming these sorts of errors are no more likely when attempting to follow BEL than when following a rule such as SUN CREAM, beliefs acquired by following BEL are safer than those acquired by following other epistemic rules, and thus more knowledge conducive.

So, although BEL is taken by Byrne to be especially knowledge conducive, it is *not* the case that it *always* produces knowledge. While following BEL always produces true second order beliefs, cases where there is a nearby possibility of believing that *one believes that p* even if one does not believe that  $p$  (i.e. making a Type III error) would prevent the true belief from being knowledge.

I will for now accept the claim that the possibility of Type III errors impacts the safety of beliefs acquired by following BEL. However, it should be noted that Type III errors arguably do not pose quite the problem Byrne takes them to. In Type I and Type II cases, the subject follows BEL and as a result ends up with a false belief. In Type III cases, one ends up with a false belief, but *not* by following BEL. Because the subject does not think conditions C obtain, it cannot be said that she reasons in accord with the rule.

Given that a plausible safety condition on knowledge will be basis-relative (i.e. you couldn't easily have believed that *p* on the basis you believe it and been wrong), the possibility of Type III errors appears to be irrelevant to the issue of safety, since they do not involving believing that *p* on the basis that C obtains.

### 3.3 A Note on the Importance of Self-Verification

It is important not to underestimate the significance of Byrne's construal of BEL as strongly self-verifying. The success of the account essentially depends upon this being the case, for the following reasons.

First, as I hope to have demonstrated in the previous section, the feature of self-verification allows the account to explain privileged, and peculiar access. The fact that BEL is self-verifying explains the privileged nature of self-knowledge by guaranteeing that self-attributions made by following the rule will be accurate (though it does not guarantee they will be knowledge). It explains peculiar access by showing why we are able to use the inference to learn of our own beliefs, but not to learn of other people's beliefs.

Second, self-verification is required for the inference to be *reliable* (in the sense that it reliably produces true beliefs). If it were the case that one could recognise *that p* and use this as a premise in inference without believing *that p*, then inferring according to the schema would not be reliable.<sup>51</sup> Indeed, this is why the inference would be unreliable if employed to learn of other people's beliefs (2018: 105).

Third, Byrne's account depends upon the feature of self-verification to explain why the inference produces *safe* beliefs. That is, it is because the inference is self-verifying that one could not easily have been wrong about what one believes when one reasons in this way. Byrne takes the feature of safety to explain why the inference is *knowledge conducive*. More specifically, on the inferentialist account, true, safe belief is taken to be *sufficient* for knowledge. It thus needs to be the case that the inference is strongly self-verifying to ensure that it produces beliefs that are not only true, but warranted such that they count as knowledge (2018: 106, 109-112).

So, in summary, Byrne's account relies upon the inference described being strongly self-verifying in order to do following: 1. explain the distinctive features of doxastic self-knowledge, 2. make the inference reliable, 3. make the inference knowledge conducive.

---

<sup>51</sup> In particular since *p* does not provide *evidential* support for the fact that one believes *p*. I will return to this issue shortly.

### 3.4 Objections and Responses

#### 3.4.1 Is the Inference Strongly Self-Verifying?

As I showed in section 1, much of Byrne's account hangs on BEL being strongly self-verifying. However, I now attempt to show that we have good reason to suspect it does not in fact have this feature.

Let us first remind ourselves why Byrne thinks BEL is self-verifying. According to his view, in order to follow BEL one must regard  $p$  as true. One cannot reason from  $p$  to *I believe that  $p$*  if one does not affirm  $p$ . Affirming  $p$  as true is treated by Byrne as identical with *believing that  $p$* . So, given that one must affirm  $p$  as true to follow BEL, one must believe *that  $p$*  in order to make the relevant inference. Thus, the conclusion of the inference, that *I believe  $p$* , is guaranteed by the fact that one makes the inference.

However, Byrne also makes the following statement: 'judging is the act that results in the state of belief' (2018: 101). In his account, the only consequence of this fact is that epistemic rules like BEL might be better phrased as 'if  $p$ , *judge that you believe that  $p$* ' (2018: 101, 2005: 102, n. 22). But taking this kind of explicit stance on the belief/judgment distinction has potentially far greater ramifications than the inferentialist account anticipates.

I take it that a charitable interpretation of Byrne will read him as claiming that to judge that  $p$  is to engage in the act of forming the belief that  $p$ . On this understanding, judgement is an occurrent event in which one affirms a propositional content, while belief is a standing state produced by doing so. If we accept this picture, then we should not merely be concerned with whether the consequent of BEL should be expressed as 'judge that you believe that  $p$ ', rather than 'believe that you believe that  $p$ '. We should also be concerned with what is happening when one recognises that  $p$  (i.e. that conditions C obtain) at the start of a chain of reasoning.

We should reflect upon what it means to take the fact that  $p$  as a premise in inference. Given that it seems to involve occurrently affirming the proposition that  $p$ , it seems entirely plausible that we should think of the inference as beginning with the judgment that  $p$ . Indeed, this seems right even if one has a standing belief state that  $p$  prior to making the inference. Since judgements can be both events of belief formation *and* events of re-affirming or revising existing beliefs, the characterisation works.

Given the fact that inference seems to involve reasoning from an occurrent judgment, Byrne's account only appears to work under the assumption that wherever one occurrently judges that  $p$ , one forms the belief that  $p$ . That is, it only works if judging that  $p$  necessarily results in believing

that  $p$ . If this is not the case, and one can occurrently judge that  $p$  without having a corresponding standing dispositional attitude towards  $p$ , then it is no longer the case that the inference is strongly self-verifying, since the conclusion is not guaranteed by making the inference.

I think it is possible, and in fact not uncommon, for subjects to make conscious, occurrent judgments which do not result in belief. Consider cases of cognitive irrationality, where one occurrently judges that  $p$ , but fails to acquire a standing commitment to  $p$  due to the influence of a recalcitrant motivating bias. Judgments should be understood as mental events which ordinarily, though not *necessarily*, result in belief.

If what I have said is right, then in order to be self-verifying the inference would need to be reformulated as ' $p$ , therefore I judge that  $p$ '. BEL would also need to be reformulated as 'if  $p$ , believe that you judge that  $p$ ' (or perhaps 'if  $p$ , judge that *you judge that*  $p$ '). But Byrne is not seeking an account of self-knowledge of occurrent judgements. Rather, the account aims to capture how we acquire self-knowledge of *belief states*. Moreover, as I will argue at length in Chapters 4 and 5, there is a more immediate way in which we know our own judgments, in virtue of the fact that they are mental actions. Consequently, we do not need an inferentialist account of judgment.

However, interestingly, if it turns out as a result of what I have suggested here that Byrne's inference is not strongly self-verifying, it is not necessarily disastrous for the account. Because judging is standardly the event of belief formation, even if judging that  $p$  doesn't *necessarily* result in belief that  $p$ , inferring from  $p$  to I believe that  $p$  is still likely to be a highly reliable process. In regard to safety, it is less clear that the inference will now be safer than inferences to conclusions about other people's mental states. But at the very least, the inference doesn't appear to be *less* safe than these kinds of inferences. Thus the account may no longer have a good explanation of privileged access, but it does seem to retain the ability to appeal to safety.

I now set aside the issue of self-verification, given that there are more pressing issues for the inferentialist transparency account. In the following section I concentrate on developing objections to Byrne's claim that external conditions can make following BEL *rational*. If the inference is irrational, it cannot be knowledge conducive.

### 3.4.2 The Dilemma for Inferentialism

The most well-known objection to Byrne's transparency account is due to Boyle, who poses the following dilemma: the inference Byrne describes must be either 'mad' or else it is epistemically redundant (2011: 230-1).

The first horn of the dilemma can be understood as follows: the inference appears mad because the fact that  $p$  has nothing to do with one's beliefs. That is, the fact that  $p$  doesn't provide evidentiary support for the conclusion of the inference, and reasoning to a conclusion about a subject matter from some premise which reveals nothing about that subject matter seems plainly irrational. For instance, it might be a fact that *it is raining in Paris*, but this fact alone doesn't suggest anything about whether or not I have the belief that *it is raining in Paris*. In this way, it is initially difficult to see how the inference Byrne describes could be understood as rational.

The second horn of the dilemma proposes a solution to the problem raised by the first. It suggests that the rationality of the inference can be preserved if the subject making the inference thinks of herself as affirming the fact that  $p$ . That is, if she treats  $p$  not merely as some fact about the world, but as a fact *she regards* as true. In this case, it would be *rational* for her to reason from the premise that  $p$  to a conclusion about her own psychology, because she understands the premise as reflecting something about her psychology.

However, this brings forth a new problem. If the subject regards the premise that  $p$  as reflecting her own affirmative stance on  $p$ , then it seems she has no need to engage in any further reasoning to learn of her attitude towards  $p$ . She already has a belief about her own attitude towards  $p$ . Thus the inference Byrne describes would be redundant, since it would not yield a new second order belief.

So, inferentialism faces a problem – seemingly, in order to prevent the inference described from being irrational, the inferentialist must take the second horn of the dilemma and concede that the subject already takes the premise that  $p$  to reflect her own attitude, thereby rendering the inference needless and defeating the account.<sup>52</sup> Accordingly, Byrne rejects that one already knows her own attitude before making the inference. As he accepts, 'to take the second horn of the dilemma would be to reject the transparency account' (2018: 123). Therefore the focus is on the charge of madness.

In response to the dilemma, Byrne points out that the inference he captures is not one that is made on the basis of evidence, as Boyle suggests. One does *not* make the inference because one takes the premise to support the conclusion.<sup>53</sup> This is treated as a point in favour of the account, since it distinguishes doxastic self-knowledge from other forms of knowledge (including of other

---

<sup>52</sup> Both reflectivism (Chapter 2) and the account I will develop (Chapters 4 and 5) can be understood as taking the second horn of Boyle's dilemma. I will argue that judgments are partly constituted by awareness that *I am judging*.

<sup>53</sup> This raises the question of what motivates the subject to make the inference, an issue which I develop in the forthcoming discussion.



people's beliefs) which depend on reasoning from evidence (2018: 123). Thus Byrne's account is committed to the idea that the premise of an inference need not provide evidentiary support for the conclusion in order for the inference to be rational.<sup>54</sup>

According to Byrne, the transparent inference is rational because it is reliable and knowledge conducive (2018: 123-4). That is, the inference is rational because it reliably produces true beliefs, and is conducive to *safe* beliefs (which Byrne takes to be sufficient for knowledge).<sup>55</sup> Because the inference is conducive to safe beliefs, it is conducive to knowledge, and because it is knowledge conducive, it is rational. Thus Byrne has an externalist story about what makes the inference rational. By this I mean that the features which supposedly make the inference rational need not be accessible to the subject.

In the following discussion I will attempt show that in spite of having the features of reliability and safety, the inference Byrne describes is nonetheless irrational. If I succeed, this will also serve as a counter to the claim that safety as Byrne understands it can be sufficient for knowledge, since an irrational inference cannot be knowledge conducive. First, I argue that reliability does not look like the kind of feature that can make an inference rational. I will then go on to outline some ways in which the doxastic schema lacks features we would standardly expect of rational inference.

My aim is to show that although one could systematically reject the features I outline as necessary conditions for rational inference, this would require conceding that the inference captured on the doxastic schema bears little resemblance to what we ordinarily accept as good, or rational, inference. Since the inference must be rational to be knowledge conducive, the doxastic schema does not look like the kind of model which could produce knowledgeable self-ascriptions.

### 3.4.2.1 Reliability

In this section I attempt to motivate the idea that reliability cannot be appealed to in explaining what makes an inference rational. I begin by outlining a case which I take to be analogous with Byrne's account of transparent self-knowledge acquisition in certain important respects.<sup>56</sup>

Justine is a fortune teller who reads cards. She sincerely believes her cards reveal the future to her. The predictions Justine makes on the basis of her cards nearly always come true. However, this is because, unbeknownst to Justine, her predictions nearly always

---

<sup>54</sup> In this way, Byrne's account is further distinguished from more traditional inferential accounts, which understand self-knowledge as being acquired by inferring from evidence.

<sup>55</sup> This is assumed. No argument is presented in support of the idea that safety is sufficient for knowledge.

<sup>56</sup> The case described can be viewed as a reworking of Bonjour's clairvoyant example (1985), which he employs as a challenge to epistemic reliabilism.

influence the way her clients behave – ordinarily, they deliberately act so as to bring about the things Justine predicts. In this way, when Justine gives a reading she ordinarily brings it about that her prediction will come true.

Although the beliefs Justine forms about her clients through card reading are reliably true, it would seem problematic to say that Justine's beliefs are rational. This is because her card readings are a poor evidential basis for her beliefs about what will happen to her clients. The cards do not in fact indicate anything about the future, and so the evidence upon which Justine's beliefs are based does not in fact support the propositional content of those beliefs.

On Byrne's view, transparent second-order beliefs about one's doxastic attitudes look worryingly like the poorly grounded beliefs Justine has about her clients' futures. Like reasoning according to the doxastic schema, the method of card reading produces beliefs that are reliably true. However, also like the card reading, reasoning according to the doxastic schema involves arriving at a true belief by reasoning from a premise which does not support this belief. We wouldn't think the fact that the card readings are reliably accurate makes the method rational. Insofar as the inferentialist method resembles Justine's card reading, we should be wary of conceiving it as rational.

However, there are important ways in which the card reading case differs from Byrne's account. Perhaps most significantly, in the card reading case, it is merely a contingent truth that Justine's method of belief acquisition is reliable. The doxastic schema on the other hand is supposed to be *necessarily* reliable, since using the method necessarily requires the conclusion to be true.<sup>57</sup>

When a subject infers from the fact that *p* to *I believe that p*, she can only make this transition on the condition that she believes that *p*. But if Justine comes to believe truly that *her client will be married within a year* by reading cards, this belief is only contingently true; it does not necessarily follow from the fact that she uses the cards to form the belief that it will be true.

Even taking into account the disanalogies between Justine's card reading and Byrne's inference, there remains a problem. When we consider what makes Justine's method irrational, it seems not to hinge on the fact that it produces beliefs that are only *contingently* true. Rather, what makes her method irrational is that she reasons from a fact that doesn't support that *p* to the conclusion that *p*. This is something shared with Byrne's inference.

The issue seems not to be with the degree of reliability, but with the fact that reliability of any kind cannot make an inference good, if the premise of that inference fails to evidentially support

---

<sup>57</sup> Though as we saw in 3.4.1 this may not in fact be the case.

the conclusion. I suggest that reliability should instead be conceived as an *upshot* of rational inference, and not something that *makes* an inference rational. Of course, Byrne would no doubt respond as he does to Boyle, by claiming that his method, unlike Justine's, is not evidence-based. Thus in the following subsections I attempt to develop some further thoughts about why an inference in which the premise does not evidentially support the conclusion is a bad one, in the sense that it seems irrational.<sup>58</sup>

### 3.4.2.2 Citing Premises

So far I have tried to motivate the idea that reliability cannot make Byrne's inference rational by appealing to an analogous case. I will now consider a further reason why the inference does not appear to be rational.

It is ordinarily characteristic of inference that I can cite premises as supporting the conclusion of the inference. For example, if I infer from the premise *I haven't been food shopping* to the conclusion *there is no guacamole in the fridge*, I can cite the fact that I haven't been shopping as evidence supporting the conclusion that there is no guacamole. But in the case of Byrne's inference, I cannot cite the fact that *p* as supporting the conclusion that *I believe that p*.

Accepting that we can make rational, knowledge conducive inferences by reasoning from premises that don't support our conclusion *means* accepting that we can acquire knowledge by reasoning from premises that we cannot *cite* as supporting our conclusion. From an empirical perspective this looks odd, but as Boyle notes, it also reveals a deeper issue. Specifically, that beliefs acquired using Byrne's inferential method appear to be unstable (2011). The thought is something like this: if I cannot identify the premise of the reasoning which lead to my belief as genuinely supporting the contents of that belief, then I end up in a position where I have no reason to hold onto that belief.<sup>59</sup> Beliefs gained by following BEL thus seem to have an inherent instability, in that they do not stand up to reflection upon one's reasons.

---

<sup>58</sup> One might ask what it is exactly for a premise to provide evidentiary support for a conclusion. While I will not give a full account of evidence here, I take it that there are different ways in which a premise can act as evidence in support of a conclusion. For instance, a premise might support a conclusion by logically entailing it. Or, that premise being true might make the conclusion highly probable. In any case, there must be some sense in which the truth status of the premise bears upon the truth status of the conclusion in order for the premise to be evidence for the conclusion. This being the case, there is no account of evidence on which the premise of the doxastic schema is evidence for its conclusion.

<sup>59</sup> In Boyle's full critique seems to develop the worry in a way that is more connected to his view of belief as a kind of active state, what he calls 'a sustained condition of finding persuasive a certain view about what is the case' (2011: 230). Aside from causing a problem in cases where one reflects on one's reasons, the fact that *p* is not a reason to believe that *I believe that p* means beliefs about one's doxastic states formed by reasoning in line with the doxastic schema are unsustainable insofar as one cannot be in a sustained state of thinking that *p* is true without having any evidence in favour of this.

One immediate problem with Boyle's claim about instability is that we do in fact appear to sustain beliefs once we have formed them *even if we do not regard those beliefs as supported by evidence*. For instance, my mother is excellent at remembering birthdays. When prompted she can accurately report the birthdays of even the most distant relatives. It is not the case that she sustains her beliefs about when these birthdays are by having continued reason-giving evidence for them (or at least what she regards as such evidence). When asked to give her reason for believing that my cousin's birthday is on the 1st June, in all likelihood she would not be able to provide one. Yet recognising that she no longer has a reason for the belief would not lead her to abandon it. I take it many other beliefs are like this. We assume that we *had* compelling reason(s) in favour of believing that *p*, and this is enough for the belief to be sustained. Moreover, it seems to be a matter of empirical fact that subjects often sustain beliefs for which they find no compelling evidence at any point, for instance recalcitrant irrational beliefs motivated by sustained desires.

However, the problem for Byrne is that reflection upon one's reasons for believing that *I believe that p*, when this second order belief has been produced by following BEL, will reveal to one *a premise which does not support one's belief* (providing one is able to accurately identify her chain of reasoning). If the subject reflects upon her second order belief only to find that its basis is something she cannot cite in support of this belief, it is natural to assume that, if she is rational, she will abandon the belief. The same applies to the example of my mother's belief – were she to recognise that her belief about my cousin's birthday was based on something that did not support its content, assuming she is rational, she would abandon the belief.

Of course, the fact that beliefs formed by following BEL are unstable in this way does not itself show them to be irrational. Nonetheless, we would not expect beliefs acquired through good, rational inference to have this problem, and Byrne has no satisfactory response to the issue.

### 3.4.2.3 Reason Giving

The fact that the premise of Byrne's inference does not support the conclusion seems to entail that one would *lack a reason* to make the inference. That is, one would lack a reason to draw the conclusion that *I believe that p*. In this subsection I consider whether this poses a problem for the rationality of the inference.

There are different things we might mean when we talk about the idea of there being a reason for belief. One thing we might mean is that something is a normative reason for belief.<sup>60</sup> That is, it is a

---

<sup>60</sup> Wedgwood (2006) suggests that rational inference involves responding to *normative reasons*.

reason for belief independent of whether I acknowledge it as doing so. So, I might not think that Charlie applying sun cream gives me a reason to believe that *he believes it is sunny*, but in a normative sense it does provide a reason. Or, we might mean that something gives a reason for belief insofar as *I regard it as giving me a reason* (insofar as I regard it as providing evidential support for the content of the belief).

The fact that Charlie is planning on walking to work might not actually provide a normative reason for believing that he believes it is sunny (perhaps, unbeknownst to me, he walks everyday), but I might nonetheless regard it as providing a reason.<sup>61</sup> For our purposes, the important thing is that on Byrne's model of inference, the subject seems not to have either kind of reason for arriving at the belief that *I believe that p*. This leads to two different issues, one concerning rationality, and one concerning motivation.

First let us consider the issue of motivation. In standard cases of inference, a subject is seemingly motivated to make the inference because she regards the premise as *giving a reason* to draw the conclusion. If I infer following SUN CREAM, I seemingly do so because I take the fact *that Charlie is applying sun cream* to give me a reason to believe that *he believes it is sunny*. The reason-giving-ness of inferences seems to stem from the fact that the premise is regarded by the subject as providing evidentiary support for the conclusion of the inference.

But it is then unclear what would motivate me to make the inference from the fact that *p* to *I believe that p*. We cannot suppose that I regard the fact that *p* as evidence for *I believe that p*. Since *p* doesn't actually give a reason for the belief, this would leave the account open to Boyle's initial madness objection. So I cannot *rationally* treat *p* as giving a reason to infer that *I believe that p*.

Now to the issue of rationality. There are two possible ways in which the fact that one does not have a reason to make Byrne's inference poses a problem for the rationality of the inference. These relate to the two different kinds of reason I outlined at the start of this subsection. The first problem is this: if a subject who does not identify the premise of her inference as giving a reason for believing its conclusion makes the inference, this looks irrational. That is, one might think that to make the inference in spite of not taking herself to have a reason to do so is to engage in bad reasoning. Of course, one could reject that the rationality of an inference is dependent upon the subject taking herself to have a reason to make it. However, this leads us to the second rationality problem.

---

<sup>61</sup> This would be an example of what I earlier labelled an 'agent's reasons'.

Byrne's inference not only fails to be reason giving in the sense that the subject identifies the premise as giving a reason to believe that conclusion. There is also no normative, agent-independent reason for believing the conclusion. The fact that  $p$  does not give a normative reason for believing that *I believe that p*. Moreover, Byrne's inference is supposed to be rational even in cases where  $p$  is false, since the inference is still meant to yield true second order beliefs about one's false beliefs. So the rationality of the inference cannot depend on the externalist condition that  $p$  is true being met, since this would preclude the possibility of it being rational to use the inference to learn of false beliefs.

What the above discussion suggests is that there must be *some* form of internalist condition on rationality. That is, what makes an inference rational must be dependent in some way on the mental states of the subject herself. More specifically, a subject must regard the premise of her inference as evidentially supporting the conclusion.<sup>62</sup> Since a subject cannot rationally regard  $p$  as supporting the conclusion that *she believes that p*, inferring in line with the doxastic schema cannot be rational.

### 3.4.2.4 Rule-Following

A further problem for the doxastic schema is that it seems to be underwritten by an implausible principle. It is widely accepted that good, rational inference involves rule-following.<sup>63</sup> That is, rational inferences are underwritten by rules which we observe in reasoning. But the rule underwriting the doxastic schema raises issues for the inferentialist account.

As we have seen, Byrne understands the doxastic schema as being underwritten by the following epistemic rule (2018: 102):

BEL if  $p$ , believe that you believe that  $p$ .

According to Byrne, BEL is a good rule because it is self-verifying, in that wherever one follows the rule, one must believe that  $p$ . Therefore one's resulting second order belief must be true. This is taken to be the case because, as Byrne puts it, 'recognizing that  $p$  is (inter alia) coming to *believe* that  $p$ ' (2018: 104). Because the rule is self-verifying it is supposedly reliable and knowledge conducive.

---

<sup>62</sup> This is a point emphasised by Boghossian, who argues that inferences must meet what he refers to as the 'Taking Condition': 'Inferring necessarily involves the thinker taking his premises to support his conclusion and drawing his conclusion because of that fact.' (2014: 5) I think some version of the taking condition must be right.

<sup>63</sup> See for instance Broome (2013), and McHugh and Way (2018).

To see why the rule is implausible, we must first note that, ordinarily, epistemic rules like BEL give us information about the conditions under which certain propositions are true (or at least likely to be true). For instance, recall SUN CREAM. This rule suggests that when certain conditions obtain, in this case Charlie applying sun cream, it is likely to be true that Charlie believes it is sunny. Likewise, if we consider Byrne's DOORBELL (if the doorbell rings, believe there is someone at the door), this rule suggests that when the doorbell rings, the right conditions obtain for making it likely that someone is at the door. We are instructed by epistemic rules to form beliefs that certain propositions are true when particular conditions obtain *because* those are the conditions under which those propositions are true (or likely to be true).

If we apply the above idea to BEL, we find that the rule suggests that when  $p$  obtains, it is true that *I believe that p*. This is somewhat alarming. It seems that BEL suggests I am omniscient; the mere fact of  $p$  being true apparently provides the conditions under which it is true that *I believe that p*. Thus it seems that Byrne's inference is underwritten by a bad rule. This represents a further way in which the doxastic schema does not meet widely held expectations for rational inference.

### 3.4.2.5 Present Tense Restriction

A further way in which the doxastic schema looks different from paradigm models of good inference is that it only works in the present tense. Inferring from *p was true* does not guarantee that *I believed that p*. Likewise, taking *p will be true* as a premise in inference does not guarantee that *I will believe that p*. So, for example, the inference from the fact that *it was raining* to the conclusion that *I believed it was raining* would not be strongly self-verifying, since to make this inference the conclusion need not be true.<sup>64</sup>

On the inferentialist account, the doxastic schema only produces safe, reliably true beliefs *because* it is strongly self-verifying. Thus, to retain the possibility of the doxastic schema being a reliably accurate model of inference, which is necessary for it to be knowledge conducive on Byrne's account, it must be restricted to the present tense.

However, rules of inference are not standardly restricted in this way. Again take the example of *modus ponens*. This kind of inference works in the same way regardless of tense. To see this more clearly, take the following case:

Rule: If Charlie is home, believe the heating is on.

---

<sup>64</sup> A similar problem, which I will not discuss here, is that the rule underwriting the inference cannot be followed in *hypothetical* reasoning. For an extended discussion of this objection see Valaris (2011).

## Inferentialism

Present tense: *if Charlie is home, believe the heating is on.*

Past tense: *If Charlie was home, believe the heating was on.*

Future tense: *If Charlie is going to be home, believe the heating will be on.*

The above example demonstrates that a good epistemic rule can be followed to secure rational inferences regardless of what tense the subject is reasoning in. Of course, it is open to Byrne to once again reject that being valid across tenses is a necessary feature of rational inference, even if it is a standard one. But this means conceding yet another way in which the doxastic schema is unlike ordinary cases of good inference.

### 3.5 Conclusion

At the start of this chapter I stated my intention to determine whether a transparency account that did not appeal to rational agency could provide a satisfying explanation of the warrant for transparent self-ascriptions of belief. That is, whether it could explain what makes transparently acquired beliefs about one's own doxastic states *count as knowledge* more successfully than the agency-based accounts. If so, I suggested that we would have good reason to think agentialism and reflectivism overestimate the importance of rational agency in doxastic self-knowledge.

Recall that reflectivism and agentialism both understood our warrant for transparent self-knowledge as deriving from our agential authority over our own beliefs. Those accounts took one's active, rational assessments (which on the reflectivist account were understood as beliefs) to be necessarily self-aware, or reflective. On both accounts, it was suggested that this reflective, self-aware nature made it the case that world-directed assessments of one's reasons non-inferentially grounded self-knowledge of belief. The accounts relied heavily on the subject possessing, and moreover deploying, a rich conceptual understanding of her own beliefs.

The inferential account considered in this chapter denies that it is necessary for the subject to have access to facts about the nature of her rational assessments in order to be entitled to transparent self-knowledge. Instead, the inferentialist account claims that one's entitlement to doxastic self-knowledge can be explained by purely external conditions. On the account, a subject is in a position to learn of her own beliefs by inferring from the premise *p* to the conclusion *I believe that p*. This is supposedly because such making such an inference guarantees the truth of the premise. It is thus regarded as 'strongly self-verifying'. The feature of being self-verifying is taken to make the inference both *reliable* and conducive to *safe* beliefs. These features are in turn taken to demonstrate that the inference is rational, and conducive to *knowledgeable* self-ascriptions of belief.



The problem for inferentialism, as Boyle points out, is that the inference it captures looks to be irrational. I developed this idea by first suggesting that the reliability of an inference cannot be appealed to in explaining what makes it rational. Reliability appears to be a fallout of rational inference, rather than itself a rationalising feature. I then identified four plausible conditions on rational inference which reasoning according to BEL fails to satisfy: that one can cite one's premise in support of one's conclusion, that the premise gives the subject a reason to believe the conclusion, that rational inferences are underwritten by good rules, and that the inference is rational across tense.

The inferentialist therefore seems forced to concede that the inference captured by the doxastic schema does not resemble standard instances of rational reasoning, and this is highly troubling. In light of the objections raised in 3.4.2, I contend that Byrne's inferential view fails to give a plausible account of what makes our transparent self-ascriptions warranted such that they count as knowledge. Doxastic self-knowledge cannot be grounded in irrational reasoning.

The most significant takeaway from the last three chapters is that the agency-based views get far closer to establishing a convincing epistemic account of transparent self-knowledge than the non-agential inferentialist account. While those accounts made broadly plausible suggestions about the relation between active rational thought and self-awareness, the inferentialist account presents a view on which transparent self-ascription looks like straightforwardly bad reasoning. In Part 2 of the thesis I will therefore seek to develop the central ideas of the agency-based views, while also avoiding the more problematic aspects of these accounts.



## Part 2: A Phenomenological Approach to Transparency

### Introduction to Part 2

In Part 1 I advanced critical analyses of three promising accounts of doxastic self-knowledge. Each of these proposed a different way of understanding how transparent self-ascriptions are warranted such that they count as knowledge.

The agentialist account advanced by Moran (Chapter 1) made the important observation that we have authority over our beliefs not just in the sense that we have special access to them, but in the sense that we are responsible for bringing them about. His account gestured towards a plausible epistemic account on which a subject's entitlement to doxastic self-knowledge derives from the fact that they are products of her rational reasons assessment. However, this account was dependent on an overly demanding form of reflective rational agency. It also left the grounding relation between deliberation and self-ascription under-explained. As a result, the account was vulnerable to the 'uniqueness' objection.

Boyle's reflectivist account (Chapter 2) addressed crucial gaps in Moran's account. It suggested that the nature of belief was active. Specifically, that belief represented a form of active *state*; a non-occurrent exercise of agency. This was supposedly due to the fact that regarding *p* as true constitutes endorsing *p* (i.e. regarding *p* as to be believed). This in turn was supposed to entail that whenever an agent believes that *p*, she also knows that *she believes that p*. The reflectivist account therefore had a kind of constitutivist story of doxastic self-ascription.

However, reflectivism's appeal to the idea of an active state proved puzzling, not least of all because it necessitated an overly complex account of epistemic failure, including repressed belief and akratic belief. The account was also flawed in its appeal to the symmetry between responsibility for belief and responsibility for intentional action. I contended that this symmetry was explainable without construing belief as itself an act. Furthermore, the account made belief into an overly reflective state, in which the subject deploys an implausibly complex conceptual understanding of that nature of belief.

Finally, I considered Byrne's inferentialist account (Chapter 3). This account gave no special role to rational agency. Instead it claimed that transparent doxastic self-knowledge was the product of a reliable, safe inference from *p* to *I believe that p*. The aim of the chapter was to determine whether the warrant for transparent self-knowledge could be explained without appeal to rational agency.

## Introduction to Part 2

I advanced an argument against the inferentialist transparency account by developing Boyle's observation that the inference captured by the doxastic schema cannot be rational. This argument consisted of five core claims: 1 - the reliability of the inference does not help make it rational, 2 - a subject who infers from a premise must be able to cite that premise in support of her conclusion, but one cannot cite *p* as supporting the conclusion *I believe p*, 3 – the fact that *p* does not give the subject a reason to infer that *she believes that p*, 4 – the doxastic schema is underwritten by a rule that is implausible, in that it suggests omniscience on the part of the subject, and 5 – the inference does not resemble ordinary cases of inference in that it can only be used in the present tense. I concluded that we shouldn't regard the inference as rational, and thus that we cannot regard it as knowledge conducive.

In Part 1, I thus established that the most promising and fully developed transparency accounts fail to adequately explain the warrant for our transparent self-ascriptions. However, we can take important lessons from these accounts as we move into Part 2. The most significant lesson is that there seems to be a connection between the fact that beliefs are the product of active thought, and our entitlement to transparent self-ascriptions. This is seemingly due to the fact that a subject understands her active assessments as *hers* in some sense. As we saw in Chapter 3, denying that rational agency plays a central role in the explaining the epistemology of doxastic self-knowledge makes it difficult to see how transparent self-ascriptions could be rational.

Other lessons we can take forward include the following: if our transparent self-knowledge of our own beliefs is informed by the fact that we self-awarely form these beliefs through active assessment, then we need to explain the precise nature of this awareness. More specifically, we need to explain how this awareness puts us in a position to knowledgeably self-ascribe belief. In light of the issues raised in Chapter 2, I think we should reject that self-ascriptions are constitutive of belief. Thus, since self-knowledge is not built into belief itself, we need to explain the method by which we make transparent doxastic self-ascriptions. That is, we need to explain how a subject goes from merely having grounds for self-knowledge, to actually possessing it.

In Chapters 4 and 5 I propose an account of self-knowledge that can explain how our rational assessments provide warrant for knowledgeable self-ascriptions of belief, as well as the method by which a subject acquires transparent self-knowledge of her beliefs. In this account, I attempt to preserve the attractive features of the agency-based accounts. However, perhaps somewhat unexpectedly, given that I have largely dismissed Byrne's inferentialism, my view also incorporates an inferential account of self-knowledge acquisition.

The account I present in the following chapters can be summarised as follows: we can explain the source of our warrant for transparent self-knowledge of beliefs by observing the nature of

*judgment*. Judgment is an occurrent, conscious mental event in which a subject presently affirms a propositional content. In other words, it is the event of a subject consciously assessing *p* as true. Judgment that *p* is *ceteris paribus* the formation of the belief that *p*. To judge is to exercise a form of non-voluntary agency: it is a kind of *mental action*.

The metaphysical status of judgment informs its phenomenology. Because judgment is a form of mental action, a subject who judges is implicitly and pre-reflectively aware that the judgment is her own. Specifically, she is aware that it is her own in the sense that it is an exercise of her agential capacity to make an assessment of *p*. In virtue of this, she understands the thought as an expression of her epistemic commitment to *p*. The subject's phenomenal awareness of her judgment that *p* constitutively grounds the knowledgeable belief that *she judges that p*. Thus, whenever the subject judges that *p*, she tacitly knows that *she judges that p*.

Because judgment is *ceteris paribus* belief formation, knowing that one judges that *p* puts one in a position to knowledgeably self-ascribe the belief that *p*, provided one is in possession of a full concept of belief (and I take it that a subject capable of self-ascribing belief will have such a concept). Since a subject who judges that *p* knows that she judges that *p*, she has warrant for self-ascribing the belief that *p* just in virtue of judging that *p*. The final question then is how she goes about making such a self-ascription.

I suggest that a subject transitions from having tacit knowledge of her own judgment that *p* to self-ascribing the belief that *p* by way of inference. Unlike the inference we find in Byrne's account, this inference is a rational one. Moreover, self-ascriptions made via the inference from *I judge that p* to *I believe that p* meet plausible conditions on knowledge, including reliability, normalcy, and safety.

## Chapter 4     The Metaphysics of Judgment

In Chapter 2 we saw that the reflectivist account is importantly informed by Boyle's metaphysical account of belief. My own account of transparent doxastic self-knowledge is primarily informed by a *phenomenological* account of *judgment*, which I understand as importantly distinct from belief. However, this account of the phenomenal character of judgment is itself connected to a particular understanding of its metaphysical status.

Moreover, my epistemological account of doxastic self-knowledge depends upon on this metaphysical account. In particular, it depends on judgment having a certain kind of metaphysical relation to belief (discussed in 4.3). For this reason, this chapter will be concerned with explicating and defending the relevant *metaphysical* account of judgment.

The following is a brief summary of the metaphysical view of judgment to be discussed at length in the present chapter:

*To judge is to consciously affirm a given propositional content. It is an occurrent mental event. Judgment is an exercise of the subject's agency, and thus represents a form of non-intentional, or non-voluntary, mental action. Judgment is ceteris paribus identical to one of the following: the formation of a belief with the same content (which may amount to the revision of an existing belief), the re-affirmation of an existing belief with the same content.*

Accordingly, the chapter will be divided into discussions of the key features identified in the above definition. In section 4.1 I detail my view of judgment as a conscious, occurrent, mental event. In 4.2 I defend the controversial idea that judging is an exercise of mental agency. In 4.3 I discuss the metaphysical relation between judgment and belief, and indicate how this has ramifications for the way doxastic knowledge, broadly speaking, can be grounded.

### 4.1     Judgment as an Occurrent Event

In this section I discuss the idea that judging is an *occurrent, conscious event of affirming a propositional content*. It is an idea I take to be broadly in-keeping with a common-sense understanding of judgment, something that can be seen most clearly by drawing a contrast with belief.

Though the terms judgment and belief are often used interchangeably, we have good reason to treat judgment and belief as separate categories of mental phenomena, albeit ones that are importantly connected to one another. On my view, while beliefs and judgments are both kinds of

epistemic attitudes<sup>65</sup> with propositional contents, judgments are occurrent events, and beliefs are non-occurrent states. This is a commonly held distinction made elsewhere by McHugh (2011), Boyle (2009), Shoemaker (2009), Setiya (2008), Shah and Velleman (2005), and Peacocke (1998, 1999, 2007). Some of these commentators also hold that judgment is a form of mental action (McHugh, Shoemaker, Shah and Velleman, and Peacocke). I discuss this more controversial claim separately in 4.2.

To see the difference between beliefs and judgments, consider the following. Suppose I have the belief that *House of the Devil is a scary film*. This belief persists over time, and it might well be possible to identify the point at which I came to have it. In the event that I change my mind it may also be possible to identify the point at which I stopped having the belief.

However, while we would regard the belief as something that exists over a period of time, it would be strange to think of it as something that is *happening*. Likewise, if I ceased having the belief it would seem inappropriate to ask me when my belief *took place*. This is the kind of language we reserve for occurrent events. It would be acceptable to ask questions such as these about my trip to the cinema, or the sunset I saw on the way home, but not, it seems, about my standing attitude towards the film. This reiterates the point observed in Chapter 2, that ordinary language suggests non-occurrent states cannot be perfectly instantiated, in that they cannot be *done* (Setiya, 2008: 38).

However, I take it that it *would* be acceptable to use the language of occurrent action in reference to my judgment that *House of the Devil is a scary film*. It makes sense to speak of judgments taking place at a time. They are, as McHugh writes, ‘individuated as dated events, not just by their contents’ (2011: 246). This also explains why we tend to refer to judgments as things we *make* and beliefs as things we *have*. To ‘make’ implies an occurrent temporal character, while ‘having’ suggests a non-occurrent condition of possession. I *make* a hand gesture, but I *have* green eyes.<sup>66</sup>

So, I accept the uncontroversial view that a belief is a standing, i.e. non-occurrent, state, while judgment is a conscious mental event of affirming a propositional content.

---

<sup>65</sup> The term ‘attitude’ is sometimes presumed to denote standing mental states, but I will use it in a broader sense which can encompass both occurrent and non-occurrent mental phenomena.

<sup>66</sup> Note that this distinction is also suggestive of judgment being active and belief being passive, in the sense that making a hand gesture is an act I perform and having green eyes is a passive condition I am in.

## 4.2 Judgment as Belief Formation

I have established that judgment is different from belief in virtue of being an occurrent, conscious, mental action. Beliefs, I have claimed, are standing states, as opposed to occurrent actions. As such they are not consciously manifest. However, judgment and belief are importantly connected to one another. In this section, I argue that judgments are standardly acts of belief formation. That is, when a subject makes a judgment, this mental action will, *ceteris paribus*, produce a belief with the same contents.

I begin by observing a description of belief due to Peacocke (1998: 88):

To make a judgment is the fundamental way to form a belief. [...] Beliefs store the contents of judgments previously made as correct contents, and these stored contents can be accessed so as to result in a conscious, subjective state of the thinker which represents the stored content as true.

In the above passage, Peacocke captures something important about the relation between beliefs and the judgments which produce them. Specifically, that beliefs *store* the contents of one's conscious judgments. It is not the case that beliefs merely imitate, or approximate, the contents of judgments. Judging that *p* does not *cause* a new attitude replicating the content of the prior mental event. Rather judging that *p* ordinarily *constitutes* forming the belief that *p*.

Jenkins (2018) provides a helpful way of expressing the idea sketched here. According to this interpretation, when a belief is acquired through judgment, the act of judging is *token identical* with the formation of the belief. Judging that *p* simply is how a subject reaches the condition, or state, of believing that *p*.

So, to form a belief, in the relevant kinds of cases, involves *securing* the content of a judgment.<sup>67</sup> But we can also note that while the act of judging that *p* standardly brings about the belief that *p*, it is possible for the two to come apart. One can have an occurrent, conscious thought which affirms the content that *p* and nonetheless fail to acquire a standing dispositional commitment to *p*. Judgment does *not* therefore entail belief. Rather, judging that *p under normal conditions*, results in having the belief that *p*.

For an example of a case in which there is an apparent disconnect between judgment and belief, we can consider Peacocke's 'university administrator' case. In this example, an administrator for a

---

<sup>67</sup> Non-relevant cases include those of tacit dispositional beliefs. For example, I might believe I locked my front door when I left my house this morning without having judged this to be the case.



university consciously judges that degrees acquired abroad are just as good as those received in her home country, only to systematically display preference for job applicants with degrees from her own place of residence (1998: 90). Here, it is plausible that the administrator makes the conscious judgment that *qualifications from abroad are as valuable as those from her home country*, while having the belief that *those from her home country are more qualified*.

The above case demonstrates that judgment doesn't guarantee belief. Cases in which a subject makes a judgment that is in conflict with a belief she holds require additional explanation. Standardly, this comes in the form of extenuating psychological conditions. For instance a recalcitrant bias may prevent a standing commitment from being formed.<sup>68</sup> This is why many cases in which a subject's belief does not match up with her occurrent judgment are identified as instances of cognitive irrationality, such as self-deception. Peacocke's university administrator case is easily construed as an example of such an anomalous psychological event. The university administrator seemingly has a recalcitrant bias against applicants from other countries, and this prevents her from acquiring the standing belief that *qualifications from abroad are as valuable as those from her home country*, in spite of the fact that she makes an occurrent judgment with that content.<sup>69</sup>

Cases of judgment which do not result in belief are thus united by a common factor. There is in each of these cases something that prevents the judgment that  $p$  that from resulting in belief that  $p$ . So, judgment is *normally* token identical with belief formation, and cases in which the two come apart are explained by anomalous psychological factors. This is all I will say on the matter for now, however in Chapter 5 I will argue that, because judgment is normally belief formation, knowing that *one judges that  $p$*  provides warrant for believing that *one believes that  $p$* . The metaphysical relation between belief and judgment will thus play an important role in my epistemology.

### 4.3 Judgment as Mental Action

In this section I discuss and defend the idea that the occurrent event of judgment is an exercise of mental agency, i.e. that it is a type of action. As previously stated, this is a view held by a number

---

<sup>68</sup> I regard biases as the product of affective states. Thus to respond to something in a motivationally biased way is to be under the influence of an affective state, such as desire or fear.

<sup>69</sup> One may cite Gendler (2008) in response to what I have claimed here. Gendler claims that attitudes such as the administrator's recalcitrant attitude are not in fact beliefs at all. Instead, implicit, affectively motivated attitudes are regarded as a different kind of state – *alief*. Unlike beliefs, aliefs are supposedly arational, in that they are not responsive to evidence. I will not attempt to deal with the worry here. However, I will note that many motivationally biased attitudes are nonetheless reason-responsive.

of philosophers. However, it is also a point of significant contention in both epistemology and ontology of mind. My aim here is not to give a definitive argument in favour of the judgment-as-agency thesis – this would be a much larger task than this thesis can accommodate, one which would require producing a set of necessary and sufficient conditions for action.<sup>70</sup> Rather, my more modest goal is to demonstrate the attractiveness and plausibility of the idea. In doing so, I lay the foundations for my later claim that judgment is partly characterised by the constitutive phenomenal feature of self-awareness.

When confronted with the judgment-as-agency thesis, one's first instinct might be to point out, rightly, that there is a very clear difference between judging and ordinary bodily action. While bodily actions are ordinarily characterised by their voluntariness, judgments are very clearly not something one can do voluntarily, in the sense that a subject cannot decide to make a judgment that  $p$  if it seems to her that  $\sim p$ . One may decide to judge at will, but one cannot make a judgment *with a specific propositional content* at will.

For our purposes, the term voluntary can be understood as essentially synonymous with intentional: an event of  $\varphi$ -ing is voluntary if one can bring it about by executing an intention to  $\varphi$ , and one cannot execute an intention to judge that  $p$ . Thus judgment is non-voluntary. For many, such as those who subscribe to a Davidsonian conception of action (1963), the fact that judgments are non-voluntary, and thus non-intentional, itself precludes the possibility of them being actions. In what follows I attempt to build a case for the idea that judging represents a form of non-voluntary action.

### 4.3.1 Responsibility for Judgments

We can begin to motivate the judgment-as-action thesis by first appealing to a symmetry between the responsibility one has for one's own judgments and that one has for intentional bodily actions. Recall that this is the same tactic employed by Boyle in his attempt to demonstrate the agentialness of *belief*. My own argument will thus bear certain similarities with Boyle's. However, I will argue that my claims do not face the same problems I raised for Boyle. That is, I will attempt to demonstrate why the responsibility we have for our own judgments is best understood as being grounded in the fact that they are themselves actions.

---

<sup>70</sup> I am sceptical as to whether it is even possible to produce such a set of conditions. Hieronymi appears to be in favour of some kind of pluralistic account of *agency*, but does not hold that non-voluntary agential events can be properly understood as actions (2009).

It is tempting to assume that responsibility, in the sense we care about when it comes to action, is dependent on the relevant event being the execution of an intention. That is, one might think intentionality is the source of agential accountability. However, I here attempt to show that the particular responsibility characteristic of action does not hinge on intentionality.

Let us begin by introducing some uncontroversial cases of agential and non-agential *bodily* events. Suppose a subject performs the intentional bodily action of pushing someone over. We would regard this as an event for which the pushing subject is fully accountable, and an uncontroversial case of action. Now suppose the subject experiences the non-intentional bodily event of *being pushed over* by someone else. We would not hold her responsible for this. In fact we view such events as being fundamentally outside of her agential control. In this way they are not events for which she can be held accountable.<sup>71</sup>

In terms of responsibility, we standardly conceive of judgment as being more like pushing someone over than getting pushed over. If a rational subject confronted with evidence clearly indicating that  $p$  judges that  $\sim p$ , under normal circumstances, we take her to be accountable for this error. In this way, we construe it as a mistake she has made, rather than something unfortunate that has happened to her. Though both being knocked over and judging that  $\sim p$  are both non-intentional events, the subject is clearly responsible for the latter in a way she is not for the former.

It will be helpful at this stage to explicitly distinguish two notions of responsibility at play in the above discussion. Namely, *agential responsibility*, and *liability*. For one to be agentially responsible for an event here simply means that the event is one's action. Being liable on the other hand means being in a position to be held *accountable*, or answerable, for an event. Ordinarily, a subject is liable for events that are her actions. However, liability for an event is not *dependent* on it being one's action.<sup>72</sup> In what follows I show why I understand liability for judgments as being grounded in agential responsibility.

Given that to be in a position to be held accountable for something can itself be considered a kind of responsibility, another way of understanding my claim here is that one form of responsibility we have for our judgments (what I term liability), is grounded in another (which I call agential responsibility). However, for the sake of clarity, I will continue to refer to the 'accountability' form

---

<sup>71</sup> This characterisation of course excludes cases of derivative responsibility, e.g. if a subject were to deliberately put herself in a situation where it was likely she would be knocked over.

<sup>72</sup> For now I leave open the question of whether one is *always* liable for one's own actions, which is not directly relevant to this discussion.

of responsibility as liability. Where I use the term responsibility I refer, unless otherwise specified, to agential responsibility.

One might initially worry that, since a subject can be liable for something without being agentially responsible for it, it isn't clear why we should accept liability for judgments as being indicative of agency. Why does being liable for judgments, insofar as we can be reasonably expected to accept criticism for them etc., require construing them as actions? After all, I could be held accountable for my dog digging up the neighbour's hydrangea, even though I am not agentially responsible for the event. Recall that this was a problem I raised for Boyle's suggestion that accountability for beliefs is symmetrical with accountability for intentional actions.

I acknowledge that there are different ways in which liability may be grounded. One way of being liable does not require agential responsibility. It involves being in a position to be merely treated *as though* one is agentially responsible, for instance by being expected to apologise or make amends. This is often the way the term is used in a legal setting. In a sense, this kind of liability might be grounded in something like agential responsibility by proxy. But a subject can also be liable because she herself is directly agentially responsible. I have claimed that this is the way in which we are liable for our judgments. In order to see why, we should consider what else, if not agency, can play the role of grounding liability for judgments.

There is one possibility we can quickly rule out, namely the idea that a subject is liable for her judgments purely because they are features of *her* mental life, and thus indicative of aspects of her character and commitments. We can dismiss this on the grounds that we wouldn't hold a subject liable for aspects of her mental life such as passing thoughts or dreams, even if these reveal or manifest features of the subject's psychology. Furthermore, liability of this sort might entail that the subject is required to *notice* when her judgments are false and to revise any resulting beliefs appropriately, but not that the subject is herself *mistaken* when she judges falsely. If she is regarded as liable for her judgments because they are exercises of her agency, then we have a straightforward answer to the question of why she is liable for them, *and* why we regard her as mistaken not just when she fails to 'make amends' for her false judgments, e.g. by revising her beliefs, but in judging falsely in the first place.

Let us consider another possibility – could liability for judgments be grounded by their connection to other mental events that are more obviously agential? For instance, could liability for judgments derive from the fact that they are produced via a temporally prior process of rational deliberation? That is, by the fact that they are produced by weighing and assessing one's reasons? This worry is not easily dealt with. It does not follow from the fact that something is produced by active processes that the event is not itself an action. That is, it may well be true that judgments

are things for which we are accountable because they are the product of active assessments, but this doesn't show that judgments are not actions. Intentionally waving my hand, for instance, can be conceived as the product of active processes on my part (forming the intention to wave my hand, executing this intention by raising my arm, and so on), but we also regard the waving as itself an action. In fact, O'Brien understands judgments as mental actions precisely *because* they are the product of rational assessments:

We think of active phenomena as originating in the subject and being under her control. [...] There will be many mental actions that are voluntary: our suppositions, assumptions, imaginings. And there will be many that are not: our judgments, doubts, denials. Nevertheless such acts of mind are subject to our rational control in a variety of ways. [...] [W]e assume that our own rational assessments direct the course of our thoughts: our thoughts are the immediate consequences of such assessments. (2007: 116).

Thus one can regard judgment as a non-agential event produced by action, or as an agential event *in virtue* of being the product of exercising one's agency. Of course, my aim is to show that judgments are themselves actions.

We should keep in mind that judgments are not merely *causally* connected to active mental processes of reasons-assessment, rather they are partly constituted by such processes. It is in the nature of a judgment to occurrently express a subject's response to her reasons. And as Moran's account brought out, this is surely a fundamental part of the explanation for why we can learn about our epistemic commitments by reflecting on our reasons – how we respond to reasons itself dictates what we judge, and consequently what we believe. As O'Brien writes, 'If our assessments as to the truth of *P* find *P* to be true, we will judge *P*, and if they find *P* to be false, we will deny *P*' (2007: 116).

In this sense, judgment consists not merely in having a conscious thought which represents a content as true, it also consists in making a conscious assessment of a subject matter, and we cannot separate these things from one another. A subject is committed to the content of her judgment because it is constituted by her own evaluative assessment of a subject matter. Given that the event of judgment just is one's actively evaluating a subject-matter, we should be inclined to think of judgments not merely as passive thoughts that fall out of mental activity, but as themselves exercises of agency.

### 4.3.2 Judgment as Aim-Directed

In the last section I argued that the responsibility we have for judgments supports thinking of them as actions. In this section I argue that the judgment-as-agency thesis is also supported the fact that judgment has a constitutive aim. Specifically, I defend the widely accepted view that judging aims at truth (Williams 1973, Peacocke 1998, Velleman 2000, Vahid 2006, Steglich-Petersen 2006, 2008, 2009, McHugh 2011). But before doing so, I want to make a clarification regarding how this topic is standardly discussed in the literature.

Commonly, it is *belief* that is described as having the aim of truth. This claim is perhaps most famously made by Williams (1973: 148). It is less common to find *judgment* explicitly referred to in this way. However, in general, talk of belief having an aim refers not to a belief state itself, but rather to *belief-forming*. As I discussed earlier, I understand the event of belief formation as, all other things being equal, token identical with judging. Hence, when I say that judgment aims at truth, I am broadly in agreement with those who claim that belief has this aim.<sup>73</sup> For the sake of consistency and clarity I will continue to refer to the event of consciously forming a belief as judgment, even where the literature cited does not use this terminology.

I take the fact that judgment is aim-directed to support the judgment-as-agency thesis insofar as actions are partially characterised by the fact that they have aims. The point then is not to show that judgments *must* be actions because they are aim-directed, but rather to show that they have something in common with uncontroversial cases of action. As a reminder, my overarching strategy in 4.3 is to build a case for thinking of judgment as action by highlighting the various ways in which judgment bears a striking resemblance to uncontroversial cases of action.

#### 4.3.2.1 Aiming at Truth

What precisely does it mean for a judgment to ‘aim at truth’? We all accept the idea that there is a standard of correctness which applies to judgments, i.e. true judgments are correct and false judgments are incorrect. A subject gets something right when she judges truly, and she gets something wrong when she judges falsely. Thus one might think there is a *metaphorical* sense in which judgments ‘aim’ at truth.<sup>74</sup> However, I contend that we should instead accept a *substantive* view of the truth-aim of belief, according to which subjects literally aim to judge truly.

What is entailed by a substantively teleological view of judgment? On a basic level, it suggests that judging is a mental action directed at a particular goal or purpose (standardly thought to be

---

<sup>73</sup> Velleman (2000) explicitly makes the distinction I have described.

<sup>74</sup> See Wedgwood (2002).

truth) which it either achieves or does not achieve. To have such an aim is an essential feature of judging - it is *constitutive* of what it is for something to count as a judgment. Thus a mental action which lacked this aim could not be considered a judgment.

Moreover, when a subject makes a judgment, it is a necessary feature of the event that she regards its content as true. That is, judging that *p* essentially requires understanding *p* as a fact about the world. Judgment proper cannot consist in merely thinking that something is *probable*, or *likely to be true*. When a subject judges that *p*, she is in that moment *epistemically committed* to *p* actually being true. The aim of judgment can therefore be spelt out more precisely: to judge that *p* only if *p* is true.

Given this particular aim, judgment is interestingly different from many other instances of aim-directed action. When I throw a ball of paper with the aim of getting it into my bin, I can perform this action without assuming it to be successful in achieving its goal. I do not aim to throw the paper into the bin *only if* the paper will land in the bin. I am also in a position to know immediately if the action fails to achieve its goal. When it comes to judging, the very act of making a judgment *presupposes* achieving the aim of truth, since judging that *p* requires conceiving of *p* as true (one may of course later discover that the judgment has failed to achieve its goal if its content turns out to be false).

If judgment proper has the constitutive aim of truth, and if a subject cannot judge without conceiving of her judgment as meeting this aim, it seems to follow that one cannot judge for purely practical reasons. Even with all the incentive in the world, I cannot judge a green object to be pink. Nor could I make a judgment purely for argument's sake, or on a lark.<sup>75</sup> To think that judgments aim at truth is to regard them as being primarily motivated by epistemic considerations, since it is these which provide the best chance of achieving the aim of judging truly. In this way, judging is unlike *supposing* or *imagining*, both of which can be primarily or wholly practically motivated.

#### 4.3.2.2 Why Accept Judgment as Aiming at Truth?

A good reason to think that judgment aims at truth is that it can help us to make sense of the standard of correctness that applies to judgment. This standard of correctness only carries normative weight because agents are in a position to be genuinely responsive to it. That is, we

---

<sup>75</sup> There is a natural way we use 'judgment' to refer to cases in which one doesn't take something to be true e.g. 'She's not sure if the defendant is guilty, but until she sees more evidence she judges that she isn't'. But these cases merely demonstrate a colloquial use of the term, where what is meant is something more like 'suppose' than 'judge'. My understanding of judgment proper excludes such cases.

## The Metaphysics of Judgment

think of an agent as being *correct* when she judges truly because she has responded appropriately to the norm that *a judgment must be true to be correct*. One way in which she can be responsive to this norm is by having truth as her goal in judging. On this understanding, judging truly is something a subject actually *tries to do*, and the aim of truth is thus a substantive one.<sup>76</sup> We can get a firmer grasp on why this should be so by considering a case where the standard of correctness cannot be applied in the same way.

There is a gag in *Parks and Recreation* that involves the character Andy playing Sudoku incorrectly. He has misunderstood the game, thinking the squares can be filled in with any old numbers. 'Sudoku is easy!' he exclaims. Now imagine, by some chance, Andy happens to get the right numbers in the right squares. Andy has given the correct answers to the puzzle. But since he wasn't aware of the rules that make these the correct answers, it doesn't seem like he could be considered agentially responsible for getting them right. He was not responding to the norms governing the game, insofar as he was not trying to create rows, columns, and squares containing one of each number from 1-9. His action would not be directed at the right aim for him to be held accountable, or indeed praiseworthy, when he correctly completes the puzzle.

When we say that a subject's judgment is correct, we mean this in more than a merely figurative or metaphorical sense. A plausible explanation for this is a judging subject, unlike Andy, literally aims at meeting a standard of correctness. Judging, on the truth-aim view, is analogous to cases in which a physical action is performed with a particular aim, for example firing an arrow with the aim of hitting a target. To hit one's 'target' in judging is to make a judgment with a true content (Sosa, 2015). Andy, having never aimed at the target relevant to playing Sudoku, cannot be thought of as having 'missed' when he fails to meet that particular standard. Or at least not in the sense that he has failed to accomplish something he was trying to do.

However, in order for a human subject to be, properly speaking, liable for her judgments, she must also be *committed* to them in the right kind of way. A subject is not liable for her suppositions or imaginings in the same way as she is for her judgments at least partly because she is not epistemically committed to their contents being true. This distinction can be brought out via an analogy with speech acts. If someone were to ask me whether it was Kurt Russell or Patrick Swayze who starred in *Point Break*, and I responded with 'I'm not sure, perhaps Russell?', I would not be held liable for this mistake in the same way that I would if I answered with 'most certainly Russell'. This is precisely because the former response does not express commitment to the falsity that Kurt Russell was in *Point Break* in the way that the latter does.

---

<sup>76</sup> Though as we have seen, this trying needn't involve a phenomenology of effort.



The truth-aim view is well equipped to capture the particular epistemic commitment a subject has towards her own judgments. This is because achieving the refined truth-aim, to judge that  $p$  only if  $p$  is true, requires a subject to judge only under the condition that the content of her judgment really is true. On this understanding, in judging that  $p$  a subject is necessarily committed to  $p$  being true.

To summarise, we should think of judgments as substantively truth-aiming as doing so provides a plausible story of the liability we have for our own judgments. First, because it conceives of human subjects as genuinely responsive to the truth-norm associated with judging. Second, because having the refined truth-aim entails epistemic commitment to the contents of judgments.<sup>77</sup>

## 4.4 Objections and Responses to the Judgment as Action Thesis

### 4.4.1 Objections to Thinking of Judgment as Aim-Directed

#### 4.4.1.1 The Over-Complication of Judgment

The notion that judging has an aim, any aim, may strike one as problematic in that it appears to give an over-complicated description of what happens when a subject makes a judgment. When I truly judge that my coffee is hot, must it be the case that this event involves me having the goal of judging truly, or rather do I judge that the coffee is hot simply because this is how I experience it? Moreover, if I did not have the aim of truth, would I not still judge the coffee to be hot? Thought of in this way, judgments seem more like reactions to stimuli than goal-orientated actions.<sup>78</sup> And indeed, this seems broadly in-keeping with how we think of many non-voluntary events.

However, reflection upon what it means to have an aim can yield a response to the above worry. One way of aiming at a particular goal is by *intending* to accomplish it. I think this is ordinarily

---

<sup>77</sup> We can extract two discrete thoughts from this section. The first is that judgment has an aim. The second is that this aim is truth. However, it should be noted that the first of these thoughts is the most important for my purposes. I am first and foremost concerned with demonstrating that judgment is an action. The agentialness of judgment consists, at least partly, in its having an aim at which it is directed. How we construe this aim is another matter. I have chosen to focus on defending the truth-aim hypothesis because it is a plausible, widely accepted option. However, it is worth being aware that there is another plausible aim that is compatible with my account. One might construe the aim of judgment as *knowledge*. According to the knowledge-aim view, a judgment aims at having a contents that is not merely true, but knowledgeable.

<sup>78</sup> This worry is *prima facie* similar to the worry about judgments lacking a phenomenology of effortfulness. However, this one does not rest on an assumption about how judgments *feel* to subjects. Instead the thought is that our judgments would be the same regardless of whether or not they had an aim – the addition of this aim to our understanding of judgment thus looks superfluous.

what we mean when we talk about aiming. A subject aims at making herself a coffee by intending to make a coffee. If my view suggested that, when one aims at judging truly, they do so with the *intention* of judging that  $p$  only if  $p$  is true, then certainly this would seem to give an over-complex, and indeed difficult to defend, account of judgment. While I can abstain from judging, it is also true that I make many judgments without first deciding to judge, and thus forming an intention to judge.

However, the account of judgment I defend does not rely on judgment being the product of an intention to capture its aim-directedness. Rather, a subject aims at truth when she judges by being responsive to evidence in such a way that manifests her aim. The distinction between the two kinds of aiming I have captured can be demonstrated through the following examples:

Suppose I am walking to work, when I see the path ahead of me is blocked by a fallen tree. Upon seeing the tree, I form the intention to take a different route into work, and change course. My action has the aim of avoiding the blocked tree and taking a different route. Moreover, I aim at avoiding the fallen tree by intending to avoid it.

But now suppose I am walking to work when a tree suddenly begins to fall into my path, and without forming the intention to avoid the falling tree I jump out of the way. Such an act might be thought of as 'automatic'. Here, I did not intend to avoid the tree, insofar as I didn't form an intention to jump out of the way before doing so, and my action was not the execution of an intention to avoid being squashed. Nonetheless, it seems appropriate to say that my jumping out of the way had the aim of avoiding getting squashed.<sup>79</sup>

So, we needn't imagine that an aim-directed event requires the kind of conceptual deployment involved in intending. Rather, aims can simply be manifest in non-intentional events. Thus the account of judgment defended here can avoid the charge of over-complexity.

### 4.4.1.2 Judgment is Non-Instrumental

Another potential problem for construing judgment as a substantively teleological event is that it appears to suggest that judgment is a *means to an end*. That is, it may be interpreted as the thought that judgment is an action carried out so as to achieve the end of having a true attitude. And yet, it is troubling to think of judgment as having this instrumental purpose. For one, it

---

<sup>79</sup> One who takes a more traditional view of action is of course likely to deny that my jumping out of the way was an action. For present purposes I will assume it is acceptable to understand the event as an action. An Anscombe-style view might regard the event as intentional if it regarded the jumping as manifesting an intention to avoid being squashed. However, I will reserve the term 'intentional' for cases in which one  $\phi$ s with an accompanying mental state of intending to  $\phi$ .

appears to jar with our standard conception of judgment as something which a subject does not just non-voluntarily, but for its own sake. A paradigmatic instrumental action would be something like going to a restaurant. It is an action performed not for its own sake, but for the purpose of achieving some other end. In the case of going to a restaurant, this end might be eating lunch. Judgments however do not seem to fall into this category.

An influential development of the objection that judging is not instrumental is due to Owens (2003), who contends that judging cannot have an aim precisely because it is not 'purposive'. On this view, judgment is thus not an appropriate action to slot into the following 'truth-aim hypothesis' (Owens, 2003: 289):

$\phi$ -ing that  $p$  aims at the truth if and only if someone who  $\phi$ s that  $p$  does so with the purpose of  $\phi$ -ing that  $p$  only if  $p$  is true.<sup>80</sup>

Owens demonstrates why the above hypothesis cannot apply to judgment by drawing a comparison with the mental act of *guessing*, which is similar to the previously mentioned mental acts of supposing and imagining. According to Owens, guessing does have the aim of truth. This is because, when a subject engages in the act of making a guess, she is able to weigh the aim of guessing truly against other aims she might have. Judgments on the other hand are only sensitive to epistemic considerations pertaining to the truth. They are not sensitive to practical considerations. Because of this, it is not possible to weigh the aim of judging truly against other aims. Thus, it cannot be the case that judgment has the aim of truth, since it is according to Owens a necessary condition for something to be an aim that it can be weighed against other aims.

Owens's point about guessing is demonstrable through the following kind of example. If I am offered a large sum of money to guess that the next election will be won by party A, even though this is in stark contrast with what my evidence suggests, I can weigh my aim of guessing truly against my aim of getting the pay out, and conceivably guess that party A will win.

In response to the criticisms raised by Owens, I first highlight a minor but noteworthy problem with his claim that guessing is a better candidate for a mental event that can aim at truth, on the basis that the aim can in this case be weighed against other aims. A guess that  $p$  is a mental act a subject carries out when her evidence is broadly in favour of  $p$ , but does not meet the standard required for judging. In this case, a guess represents what the subject takes to be most likely on

---

<sup>80</sup> We should note that for Owens, the notion of *having purpose* appears to be bound up with *intentionality*. In line with what I have previously argued, I will assume that an action can have purpose without being intentional.

the basis of her evidence. But on this understanding, it is unclear why Owens takes the aim of truth to be weighable against other aims in guessing but not in judging.

Of course, a subject could certainly carry out the *speech act* (internally or out loud) of 'guessing' against her evidence. But this act would seem not to constitute guessing in the sense Owens seeks to capture. A subject who is offered money to guess against her evidence could perhaps verbally endorse a proposition she regarded as unlikely, but this speech act could not be said to constitute a guess if it was not motivated by a concern for the evidence. Owens's objection thus seems to apply as much to the notion of guessing as truth-aiming as it does to the idea of judgment as truth-aiming.

Leaving aside this minor worry, I now consider a more substantive response to Owens's objection. As Steglich-Petersen (2006) notes, there are two possible ways of responding to Owens. One route is to deny that an aim must be weighable against other aims. The other is to claim that the aim of truth in judging *is* weighable against other aims. I will here pursue the latter form of response.

As we saw, Owens argues that any aim a subject has will exist within a broader overall network of goals. For a subject to pursue a goal is for her to pursue it in such a way that accommodates, or is most likely to facilitate the achievement of, her other goals. It is in this sense the goal is weighable against her other aims. If a subject recognises that she can accomplish a particular aim by  $\phi$ ing, but also recognises that doing so will impede her ability to accomplish another aim, then she has reason to consider ways of achieving her present aim other than  $\phi$ ing.

For instance, suppose I have the aim of getting to work, but I also have the further aim of reducing my carbon emissions. I might recognise that one way of achieving my goal of getting to work is by driving there. However, I may also recognise that driving will compromise my goal of reducing my carbon emissions. I might consequently consider other means of accomplishing my aim of getting to work, such as cycling or catching the train, in light of my goal to reduce my carbon emissions.

Owens's objection appeals to the idea that when a subject comes to believe something, an event I am understanding as making a judgment that  $p$ , this is not something she does by weighing her aim of truth against her other aims. McHugh demonstrates Owens's point through the following example: 'you cannot decide that the evidence you have for  $p$  is good enough for you to believe it, given that it's important to take a view and you can't afford to devote further cognitive resources to inquiring into whether  $p$ ' (2011: 266). It supposedly follows from this that judging cannot have an aim.

However, as McHugh (2011) suggests, judging *does* in fact take into account aims other than truth. It is simply the case that the aim of truth plays a particularly significant role in motivating judgment, partly in virtue of the fact that it is a constitutive aim of the act. That is, the fact that truth is a constitutive aim of judgment prevents it from being possible to completely disregard one's evidence in judging. The aim of truth, while it may be weighable against other aims, will always motivate judgment.

Given that truth is the constitutive goal of judging, you cannot judge in disregard of the evidence, and so you cannot judge merely as an outcome of deliberating about whether, given your overall goals and projects, to judge that *p*. This is compatible with the goal being a genuine goal. It is not that the putative goal has some extraordinary properties; it is rather that if you are not pursuing it in the right way then you don't count as judging. (McHugh, 2011: 267)

O'Brien (2007) likewise appears to regard judgment as an event which can be practically motivated:

Also in play [during judgment] will be the subject's weighting of her reasons, *the practical consequences of moving to judgment or not*, the subject's cognitive tendencies – on whether she tends to judge rashly with scant reason or is cautious about judging, requiring a surfeit of evidential reassurance. (O'Brien, 2007: 116, emphasis my own)

The crucial point is this – a mental event can only be called a judgment if it involves the subject pursuing the aim of truth. However, this does not mean that she cannot be moved by other aims, including pragmatic considerations. In footnote 19 I suggested that a subject will ordinarily be more likely to suspend judgment in high risk cases, i.e. in cases where the cost of believing falsely is especially great. This demonstrates one way in which judgments can be motivated by pragmatic concerns.

However, I contend that while judgments may be motivated by pragmatic concerns, truth must be conceived as a *constitutive* aim of judgment, for the following reason: a thought can only count as a judgment if the subject regards the content of that thought as true. Judgment is, by definition, the event of affirming a proposition. If it were possible for a subject to judge without aiming at truth (that is, to judge motivated solely by non-evidential considerations), then it would be possible for her to make a judgment the content of which she had no epistemic commitment to. Thus, given that judgments necessarily commit the subject to their content, and given that this is only possible if the judging subject aims at truth, it follows that the event of judgment must have the aim of truth.

#### 4.4.2 Phenomenology Based Objections to the Judgment as Action Thesis

In this section I defend the judgment as agency thesis from two objections. These represent alternative ways of developing the thought that judgments lack the appropriate phenomenology to count as action. I take this thought to partly motivate some of those who reject the judgment as agency thesis, most notably Strawson (2003).

Strawson's own view of mental action is outlined in the following passage (2003: 231).

[T]he role of genuine action in thought is at best indirect. It is entirely prefatory, it is essentially—merely— catalytic. For what actually happens, when one wants to think about some issue or work something out? If the issue is a difficult one, then there may well be a distinct, and distinctive, phenomenon of setting one's mind at the problem, and this phenomenon, I think, may well be a matter of action. It may involve rapidly and silently imaging key words or sentences to oneself, rehearsing inferential transitions, refreshing images of a scene, and these acts of priming, which may be regularly repeated once things are under way, are likely to be fully fledged actions.

As is evident from this description, Strawson's construal of mental action excludes judgment because, on the view he advances, mental agency consists only in the management, prompting, and directing of such thoughts.<sup>81</sup>

What motivates Strawson's view that only this limited range of mental phenomena could be agential? Like me, Strawson construes judgment as non-intentional, however he adheres to the standard view that an action must be something a subject intends to do, at least under some description.<sup>82</sup> As a result of this commitment, he is led to reject the notion that a judgment can be a kind of mental action. Of course, I have claimed that there is reason to suppose that judgments are agential even if they are non-intentional, and one might think that this in itself serves as a response to Strawson. But closer consideration of why Strawson holds judgments to be non-intentional reveals a potential problem for my view.

It appears that Strawson is partly motivated to treat mental events such as judgment as non-intentional (and thus as non-agential) by an observation about their phenomenology. Specifically, that it *feels* to subjects that their judgments occur without them doing anything. He writes, 'if we

---

<sup>81</sup> In this sense, it seems he is willing to accept only something like the form of mental agency Hieronymi refers to as 'managerial' or 'manipulative control' (2009: 140). Mele (2009) seems to take a similar view to Strawson about the kinds of things that can count as mental actions, though he thinks there is more action involved in bring about mental events like judgments than Strawson allows for.

<sup>82</sup> This is the picture advanced by Davidson (1980: 50).

consider things plainly, we find, I think, that most of our thoughts – our thought-contents – *just happen*’ (2009: 228). A natural way to read this is that we don’t feel like we are doing anything when we experience mental events like judgment. If this is the case, then judgments are not phenomenologically distinguishable from non-agential events. This is a problem *regardless of whether we construe action as necessarily intentional*; it is reasonable to expect that any kind of genuine action will feel different for the agent performing it from things that happen to her.

In the next two sub-sections I consider two ways of developing a phenomenology based objection to the judgment as action thesis. The first challenge goes as follows: (1) Judgments do not feel effortful for the subject experiencing them. Actions are necessarily phenomenologically effortful. Therefore, judgments cannot be actions. The second challenge is this: (2) Subjects do not experience their own judgments as things over which they have agential control. When a subject performs an action, she necessarily experiences it as something over which she has agential control. Therefore, judgments cannot be actions.

I claim that the first objection, that judgments cannot be actions because they are not phenomenologically effortful, should be dismissed on the grounds that effortfulness is a bad criterion for action. I do so by appealing to uncontroversial cases of bodily action that intuitively lack the phenomenal feature of effortfulness. I also consider other mental events that seem to lack the feature of effortfulness, but are plausibly agential. For instance, some cases of imagining.

The second objection, that subjects do not experience their judgments as being under their control, poses a greater threat to the judgment as agency thesis. However, I argue that this objection can be addressed by considering what it would mean for a subject to judge *without* having an accompanying feeling of being in control of that judgment. I here appeal to the psychological disorder of ‘thought-insertion’ to help illustrate why such a mental event would be one from which the subject was *epistemically alienated*. I argue that thoughts which are alienated in this way cannot constitute judgments. I conclude that the phenomenology of judgment not only permits construing it as an action, but in fact supports such a picture.

#### 4.4.2.1 Effortfulness

As stated, one interpretation of the claim that judgments ‘just happen’ is that they are not effortful for the subject who performs them. To be clear, effort in this context is quite different from *trying*.<sup>83</sup> The mere fact that a subject tries to do something does not mean that this event

---

<sup>83</sup> The notion of trying and its connection to action has been discussed by Hornsby (1980), O’Shaughnessy (1973, 1980), and Cleveland (1997), among others.

will feel effortful for her. Similarly, effortfulness in the sense I use it should not be understood as equivalent to *exertion*, i.e. expending energy. Effortfulness here captures a purely phenomenal feature, one which may or may not accompany instances of trying or exertion.

If the fact that judgment is not effortful in this sense is to pose a challenge to the idea of judgment as action, then it must be the case that actions are always phenomenologically effortful. In other words, effortfulness must be understood as a necessary condition for action. If one takes this to be the case, then they will likely adopt one of the following positions:

1. All actions must be (phenomenologically) effortful events. Effortfulness is never a feature of non-intentional events. Therefore non-intentional events cannot count as actions.
2. Non-intentional events may or may not be effortful, but if they lack the feature of effortfulness then they cannot count as actions.<sup>84</sup>

Evidently, given the assumption that judgments are not phenomenologically effortful, both of these positions rule out judgments as actions.<sup>85</sup> But it isn't at all obvious why an event should require a phenomenology of effort in order to count as an action. The amount of effort felt in performing an action can of course vary greatly, depending on the nature of the action and the subject performing it. Some bodily actions, for example, require large amounts of effort, others very little. Intentionally diverting my eyes from an unpleasant sight arguably involves no effort, whereas lifting a heavy object involves a great deal.

Another example of a form of bodily action which seems not to feel effortful for the agent, is that of highly skilled or practised action. A skilled sportsperson or craftsperson, for example, may be able to perform all manner of both basic and technically precise actions without these actions being *phenomenologically* effortful. Likewise, experienced drivers may perform intentional actions while on the road that are well-practised enough to feel literally effortless.

Mental events other than judgment may also be agential without feeling effortful. For instance, suppose for some reason I decide to imagine an elephant.<sup>86</sup> The mental image of an elephant

---

<sup>84</sup> The idea of a phenomenologically effortful but non-intentional event may sound puzzling. However, I take it that the notion is not totally unintuitive. For example, effortfulness might be present in the phenomenology of raising one's hand to bat away an object flying towards one's face, even if this event is non-intentional.

<sup>85</sup> A natural question is whether this is true of all judgments. I have claimed that having a phenomenology of effortfulness is not a necessary feature of action, and for the purposes of responding to the worry at hand this is enough. But it seems at least *prima facie* plausible that some judgments will involve a feeling of effort, for instance those that involve affirming a content one desires to be false.

<sup>86</sup> It should be noted that Strawson also rejects that imagining can be done intentionally, and thus denies it can be an action (Strawson, 2003: 239–42). This is a minority view however, and others who defend the



might come so readily to my mind that it does not have any accompanying feeling of effort. Nonetheless, it is not a stretch to think that imagining the elephant is something for which I am agentially responsible. In a similar example, I might plausibly exercise agential control over my daydreams, including the content of these thoughts, without feeling this to be an effortful activity.

Having a phenomenology of effortfulness thus seems unlikely to be a necessary feature of action, in either the mental or bodily case. For this reason, the fact that judgments are not phenomenologically effortful does not give us reason to reject the idea of judgment as action.

#### 4.4.2.2 Agential control

In this subsection I consider a different kind of phenomenological response to the judgment as agency thesis, which I take to pose a greater challenge than the effortfulness objection. According to this alternative objection, judgments lack the phenomenology of action since subjects do not experience their own judgments as things over which they exercise agential *control*, and an event a subject does not feel in control of cannot be an action.

I argue that we can respond to the agential control objection by reflecting upon what it would mean for a subject to experience a judgment-like mental event *without* feeling herself to be in control of her this event. I will claim that we must experience our judgments as things over which we exercise agential control in order for them to count as judgments at all. In this way, a phenomenology of agential control is a constitutive feature of judgment. I begin by appealing to the psychological phenomenon of ‘thought insertion’ as a means of motivating this idea.

Thought-insertion can be briefly characterised in the following way – a delusional subject experiences thoughts which she regards as being ‘inserted’ into her mind by something or someone external to herself. In the following well-known case from Mellor (1970) a patient details their experience of thought-insertion. What they describe is typical of many others with the delusion.

‘I look out the window and I think that the garden looks nice and the grass looks cool, but the thoughts of [television presenter] Eamonn Andrews come into my mind. There are no other thoughts there, only his ... He treats my mind like a screen and flashes thoughts onto it like you flash a picture.’

---

idea that judgments are non-agential generally take a more inclusive view of what might constitute mental agency.

Plausibly, the delusional subject in cases such as this is motivated to attribute the relevant mental events to an external source (usually another agent) by the fact that she does not feel *in control* of these thoughts. One of the most well-known examinations of the phenomenon of thought insertion is due to Graham and Stephens (2000), who refer to it as an instance of what they term ‘alienated self-consciousness’. They claim that thinking of oneself as agentially responsible for one’s own thoughts is central to the nature of self-consciousness. This stands in contrast to the view that self-consciousness consists merely in being aware of the subjectivity of one’s own thoughts. Hence, on their account, thought insertion cases, in lacking the phenomenology of an agential control, represent failures or breakdowns of self-consciousness.

Of course, a subject could nonetheless conceive of an alienated thought as *hers* in some sense, just as she could conceive of a sensation of pain as hers without feeling agentially responsible for it. Consider the following distinction from Coliva (2002: 31)<sup>87</sup>:

1. ‘[T]he ownership of a thought’
2. ‘[T]he sense of ownership (see Campbell 1999, 617) or of authorship or of agency over one’s thoughts’

To be clear, the sense of ownership (2) is not intended to refer to some form of awareness of (1). Rather, the term ‘ownership’ refers to substantively different things in (1) and (2). The ‘ownership of a thought’ (1) consists in being aware that one has a thought. Thus it represents a relatively thin form of ownership, of the sort a subject might have in regard to her pains, etc. The ‘sense of ownership’ (2) on the other hand refers to the feeling a subject has of exercising agential control over her own thoughts. We can thus think of ownership of this kind as *strong ownership*.

As Coliva notes, in cases of thought insertion it seems the subjects retain ownership of their alienated thoughts (1), while lacking the sense of ownership of these thoughts (2). A judging subject is aware not merely that she is experiencing a judgment, but also that she herself is the author or agent of that judgment. However, as Coliva points out, not all thoughts a subject has that lack (2) should be considered problematically alienated (2002: 32):

In favor of the independence between the relation of ownership and that of the sense of ownership, consider the case of episodic thoughts. These thoughts do not seem to play any further role in our cognitive lives; they do not connect with other thoughts, emotions, and actions, and they do not seem to be the product of antecedent thoughts

---

<sup>87</sup> It should be noted that while I regard Coliva’s distinction as a helpful tool in spelling out my view, Coliva herself explicitly rejects that judgments are phenomenologically distinguishable from other conscious thoughts, such as wishes (2008).

of ours. Yet, nobody would deny that they are the subject's own thoughts, although, while having them, she does not really experience herself as their owner/author/agent.

It may not be immediately obvious from the above passage what Coliva is referring to when she talks about 'episodic' thoughts. To clarify, the thoughts Coliva is describing would perhaps better understood as passing thoughts, such as those we might have when engaged in stream-of-consciousness-type thinking. Thoughts of this kind do not have a phenomenology of authorship – subjects do not feel in control of them. But I think it is important that we conceive of *judgments* as distinct from such thoughts.

What is key here is that judgments do play an important role in our cognitive lives, *one which they could not play were we not to experience them as exercises of our own agency*. They connect with and are sensitive to our wider network of attitudes, in a way that neither passing thoughts, nor the thoughts experienced by thought-insertion patients can. Such thoughts are unlikely to be representative of epistemic commitments in the way that judgments are. They do not require that one *affirms* their propositional content. It is because judgments play a significant role in our mental lives that lacking a sense of agential control over them would result in problematic forms of alienation, thus preventing them from having the status of judgment-proper.

Might there however be other forms of alienated thought that cause a problem for my account of judgment? According to Cassam (2011), obsessive thoughts pose a challenge to the idea that judgments are actions, since he takes them to be a form of 'passive' judgment. We can understand obsessive thought according to the following definition due to Frankfurt (1977): 'Obsessive thoughts are thoughts where one finds oneself constantly and unwillingly thinking something'. Obsessive thoughts are distinct from passing thoughts in that they seem to, at least some of the time, involve the subject occurrently affirming their contents in such a way that commits them to these contents.

According to Cassam, obsessive thoughts are judgments that are '(i) not necessarily responsive to reason, and (ii) states from which one can distance or dissociate oneself' (2011: 3). I take it that a mental event which satisfied Cassam's second condition *might* be one that lacked the phenomenal feature of strong ownership, depending on the manner in which one dissociates from them. If one dissociates later on, after thought itself has occurred, then it is possible that the thought itself nonetheless had the phenomenal feature of strong ownership. However, if the subject is distanced or dissociated from the thought during the event itself, then such a thought lacks the feature of strong ownership.

Obsessive thoughts pose a problem for my account only if they are conceived as mental events in which a subject occurrently affirms a content such that she commits to it, while also occurrently dissociating from the event. Such an event, if possible, would play the epistemic role of judgment while also lacking the phenomenal feature of strong ownership. However, I do not think we need to accept this account of obsessive thought. To see why obsessive thought doesn't pose a threat to my view, I cite O'Brien's response to Cassam.

According to O'Brien (2013), we can distinguish two different kinds of obsessive thought. The first kind of obsessive thought is formed on the basis of faulty or bad reasons (2013: 101). Such a thought could reasonably be treated as a judgment, given that it constitutes an active assessment on the part of the subject. It thus represents a genuine epistemic commitment. This may be a globally irrational one from which the subject later distances herself, but the key point for our purposes is that there is nothing to suggest such a judgment couldn't have the phenomenal feature of strong ownership, at least in the event of judgement.

The other kind of obsessive thought is one which is formed on the basis of *no* reasons. Such thoughts are classed by O'Brien as 'mere reactions or automatic responses' (2013: 101), which the subject may also dissociate from, and should not be regarded as epistemic commitments. They would be closer to the passing thoughts Coliva describes. I am happy to accept that thoughts of this kind do not have the feature of strong ownership. However, since they do not involve epistemic commitment to their contents, such thoughts are not judgments, and as such couldn't play a role in grounding doxastic self-knowledge.

So, obsessive thoughts in which a subject affirms and occurrently commits to a given content are judgments in my sense. These thoughts will plausibly have the phenomenal feature of strong ownership, even if the subject later dissociates from them. However, obsessive thoughts which occur in a similar manner to non-obsessive passing thoughts do not entail epistemic commitment on the part of the subject, in that she is not occurrently committed to their contents being true. Thus such thoughts cannot be considered judgments, and it is not a problem if they lack the phenomenal feature of strong ownership.

## 4.5 Conclusion

In this chapter I have defended the metaphysical view that judgment is a form of mental action. I began with the uncontroversial claim that judgment is an occurrent event, which is normally the event of belief formation. I then attempted to build a case for thinking of judgment as an exercise of agency by demonstrating ways in which it has features characteristic of action. First, I attempted to show that our liability, or accountability, for judgments is explained by

understanding them as actions we perform. I then argued that judgments have the structure of agential events in that they are aim-directed.

In section 4.4. I considered two sets of objections to the judgment as action thesis. The first of these objected to the idea that judgment is aim directed. I considered whether thinking of judgment as aim-directed over-complicated the event. However, I argued that since one needn't have an *intention* in order to have an aim, we don't need to worry about the event requiring a high degree of conceptually sophisticated thinking. I then considered whether it is a problem for the view that judgment appears not to be instrumental, in that the aim of truth cannot be weighed against other aims. I defended the claim that, while truth is its constitutive aim, judgments are also sensitive to practical considerations.

The second set of objections developed the idea that judgment lacks the appropriate phenomenology to count as action. The first phenomenology-based objection advanced the thought that judgment, being phenomenologically effortless, cannot be action. In order for this objection to pose a real threat, effortfulness needed to be a necessary feature of action. However, I argued against this by appealing to phenomenologically effortless events that we would uncontroversially regard as actions.

I then considered a more compelling version of the phenomenological worry. This was based on the thought that actions feel to the subject performing them like events that are under her control. If judgments, being non-voluntary events, feel to the subject like events over which she does not exercise control, then it looks as though they do not have the phenomenology we would expect of actions. However, in response to this worry I argued that thoughts which lack the phenomenology of agential control, what we can identify as 'strong ownership', do not seem to involve the kind of epistemic commitment characteristic of judgment. I concluded that this indicates that strong ownership is likely to be a feature of thoughts that do involve epistemic commitment. Thus, given that strong ownership is plausibly a feature of actions, it follows that we can think of judgments as having the phenomenology of action.

My discussion of the phenomenology-based objections can be viewed as having had two key aims. The first, more modest aim, was to show that the phenomenology of judgment does not exclude the possibility that judgment is a form of mental action. The second, more ambitious aim, was to show that the phenomenology of judgment actually *supports* conceiving of it as action, rather than merely permitting this. Though there is doubtless more work to be done on this topic, I will presume for the time being that both aims have met. Accordingly, I will proceed under the assumption that judgments have a phenomenology of strong ownership.

## The Metaphysics of Judgment

In the next chapter I develop a more detailed account of the phenomenology of judgment. In this account, I build upon the idea defended in 4.4.2 that judging has a phenomenology of strong ownership. This strong ownership, I will argue, is best understood as a form of ‘mineness’, a notion developed in recent work by Dan Zahavi, and Uriah Kriegel (Kriegel 2003, 2009; Zahavi 2000, 2005, 2011, and 2014; co-authored 2015).

I go on to claim that ‘judgment-mineness’, i.e. the phenomenal awareness one has of one’s own judgments, constitutively grounds the knowledgeable belief that one judges that *p*. This means that whenever one judges that *p*, one also knows that *one judges that p*. Since judging is *ceteris paribus* belief formation, knowledge that one judges that *p* warrants one in believing that *one believes that p*. Thus, whenever one judges that *p*, one has entitlement to knowledge that *one believes that p*. The phenomenology of judgment thus offers an explanation of the source of warrant for transparent self-knowledge. Finally, I conclude the chapter, and the thesis as a whole, by presenting an inferential account of the method by which one transparently self-ascribes belief.

## Chapter 5      Phenomenal Awareness of Judgments

In the previous section, I gave an account of judgment according to which it is a non-voluntary mental action. I argued that it aims at truth, and is standardly a way of forming beliefs. In this section I develop an account of the phenomenology of judgment. This is a subject I touched upon in 4.4.2, when I dealt with the objection that judgment cannot be agential due to the fact that it is not phenomenologically effortful. There I conceded the point that most, if not all, judgments lacked this feature, but claimed that an event need not be effortful in order to count as either a voluntary or non-voluntary action.

Here, I claim that a constitutive feature of the phenomenology of judgment is that the judge has a particular kind of phenomenal awareness of the event. Specifically, that she is aware of the judgment as being *hers*. I identify this awareness as the feature of ‘mineness’, a term I derive from recent literature.

### 5.1      Consciousness and self-awareness

It is often thought that conscious thoughts, including judgment, necessarily have a particular phenomenal feel to them, arising from the fact that there is a subject who is having the experience. This ‘feel’, what we might refer to as the ‘phenomenal character’ of an experience, has been identified in the wider literature on consciousness as essential to understanding what it means to have a conscious experience. Nagel famously refers to the ‘subjective character of experience’ (1974), the way an experience feels for the subject having it, claiming that consciousness minimally consists in this particular phenomenal character: ‘the fact that an organism has conscious experience at all means, basically, that there is something it is like to be that organism’ (1974: 436). I here remain neutral on the question of whether a phenomenal ‘feel’ is required for an experience to count as conscious. However, I contend that having a particular phenomenal feel *is* an essential feature of the conscious act of judging. On my view, this phenomenal quality is, at least in part, a form of self-awareness.

We should at this point note two commonplace assumptions about awareness of one’s own conscious episodes. Often, a subject’s being aware of something about her mental life is regarded as a separate mental state from its object. For example, if a subject were to be aware of her desire that *p*, we would consider this awareness a distinct state from the desire itself. On this kind of view, being aware that one desires that *p* might consist in the belief *that one desires that p*.

Moreover, we would standardly take this kind of awareness to be the product of a reflective process. That is, we do not consider the subject's awareness of the first order attitude to be part and parcel of what it is for her to have that attitude. To use my previous example, it would not usually be thought that the belief that *one desires that p* always accompanies the desire that *p*. Instead, it is generally thought that some process of introspection is required to achieve awareness of one's own attitudes. This introspective process can be described in terms of *attention* – when a subject becomes aware of her own attitudes, she does so by directing her attention 'inwards' towards her own psychology.

The standard assumptions can be applied to judgment in the following way: first, that a subject's awareness of her own judgment is a separate attitude from the judgment itself. Second, that there is always an extra step involved in her becoming self-aware, traditionally some form of introspection. I intend to reject both of these assumptions. In the following section I outline the details of my positive account.

## 5.2 Judgment-Mineness

The claim that a special subjective character is an implicit feature of the phenomenology of conscious experiences is notably a central point of agreement for several 20<sup>th</sup> century thinkers, including Satre (1943), Merleau-Ponty (1945), and Husserl (1959). It is a thought which has resurfaced in the work of several later philosophers, and has gained traction in recent discussions of self-consciousness. Among the contemporary philosophers who explicitly express commitment to the Phenomenological approach are Levine (2001), Zahavi, and Kriegel (Kriegel 2003, 2009; Zahavi 2000, 2005, 2011, and 2014; co-authored 2015). It is important to reiterate that my own account will not be committed to the view that conscious experiences are *always* first-personal, which Zahavi and Kriegel identify as the 'stronger claim' (SC) of Phenomenology (2015: 37). That is, I do not commit to the view that every conscious experience is partly constituted by some form of self-consciousness. I thus leave open the possibility that a subject can have a phenomenally conscious experience *without* being aware of said experience.

What I will claim is that a particular kind of awareness is necessarily part of the phenomenology of judgment. There are of course various possible ways of understanding the nature of phenomenal awareness. Guillot (2017) has recently done work to clarify the various available positions. I will here outline the distinctions she makes with a view to elucidating my own stance. Guillot identifies three ways of capturing the apparent 'subjective character' of experiences (that is, their constitutive first-personal quality). She refers to these as 'for-me-ness', 'me-ness', and



'mineness'.<sup>88</sup> In the following descriptions of these I will use the example of a subject having the experience of seeing a book.

**For-me-ness: 'the object of awareness is the experience itself'.**

The special subjective character of experiences on this view refers to the *privileged awareness* a subject has of her own experience. I.e. it refers to the fact that the experience is exclusively for that subject, and no one else may access it in the same way. This is because the subject is aware of the experience simply in virtue of having it. On this understanding, the subject is just the thing which the experience is 'for' or 'given to'. Her awareness is of her experience, as opposed to merely the object of the experience, and does not involve being aware of herself. As such, it is not strictly speaking *self-awareness*.

So, on this view, the subject seeing a book is aware of her experience of the book.

**Me-ness: 'places the focus of awareness on the subject herself, rather than (just) her experience'.**

On this view, when a subject has a phenomenally conscious experience she is aware of not just the experience, but of herself. Such awareness is standardly conceived as a *constitutive* feature of conscious experience (Kant, Fichte, Husserl, arguably Descartes). The idea seems to be something like, whenever a subject has an experience, it is a feature of this experience that she is implicitly aware of herself the subject.

On this view the subject seeing a book is aware of at least two things: her experience of the book, and herself.

**Mineness: a phenomenal awareness that my experiences are mine.**

On this understanding, the subject is aware of the experience *belonging* to her in some sense. I am aware that my experiences are mine.

This interpretation of the subjective character of conscious experiences is more nuanced than the other two. To return to our running example, the subject who has a conscious experience of seeing a book would be aware of the following things: her experience of the book, herself, and the relation of ownership between these things.

---

<sup>88</sup> Importantly, Guillot notes that a number contemporary accounts are flawed in that they use these three notions interchangeably, thus confusing their views.

My view will express a form of what Guillot identifies as *mineness*. One might expect this to be a difficult view to defend, given the complexity of this particular version of subjective character compared to the other options on the table.

For now I want to concentrate on the idea that we have a form of phenomenal awareness of our judgments best characterised as what Guillot calls 'mineness'. Recall that in the previous chapter I defended the idea that judgment has the phenomenology of action in that it involves the subject regarding her judgment as something over which she exercises agential control. I there employed a distinction due to Coliva between two kinds of thought ownership. The first, the weak kind, consisted only in regarding a thought as a feature of one's own mental life. The second, strong form of ownership involved regarding a thought as a product of one's agency, or authorship. I argued that judgments are thoughts towards which subjects feel a strong form of ownership. I claimed that such a form of ownership was required for judgments to epistemically commit a subject to their contents. This was demonstrated by comparing judgment to thoughts over which one feels only a weak form of ownership: inserted thoughts, and passing thoughts.

I take it that both for-me-ness and me-ness fail to capture the strong kind of ownership distinctive of judgment.<sup>89</sup> This is because neither involve the subject being aware of her thought as something over which she exercises agential control, in that they do not involve the subject being aware of *any* particular relation between herself and her thought. Rather, for-me-ness simply involves the subject being aware of her *experience*, and not herself or the relation between herself and her experience. Me-ness involves awareness of her *experience and herself*, but not of the relation between these things. Mineness is the only conception of phenomenal awareness which construes the subject as being aware of her experience (in this case a thought), herself, *and* the relation of ownership between these things.

Though I am interested in the mineness associated with judging, and neutral on whether it is a feature of other conscious events, it is nonetheless worth noting that appealing to mineness is an intuitive way of distinguishing what makes first-person experiences in general phenomenologically distinctive. Moreover, it is reasonable to expect there should be a common explanation for the nature of conscious experiences across the board. Thus it is of benefit to my view if it helps to explain conscious episodes other than judgment.

Consider the example of pain. When I have an experience of being in pain, there is an exclusivity to this experience – no one else can experience *my* pain in the same way that I do (though they

---

<sup>89</sup> They also fail to capture the weak form, since this also involves awareness of a relation of ownership between the subject and the thought, albeit a weakened form of ownership.

may have phenomenologically similar experiences of their own), and they cannot be aware of my pain purely in virtue of me having it (they rely on cues from me, e.g. visible, audible). However, it is also the case that I am *aware* of this exclusivity, and in being aware of the pain as exclusively given to me, I am aware of it as *my* pain. Construing human agents as being aware of the exclusivity of their experiences would thus be congruous with what is commonly referred to as ‘Immunity to Error through Misidentification’, the notion that a subject cannot be in error by mistaking who the subject having her experience is.<sup>90</sup>

The kind of mineness constitutive of the subjective phenomenal character of judgments takes a distinctive form. When a subject makes a conscious judgment that *p*, in being implicitly and pre-reflectively aware of it as *hers*, she is also aware of it as representing her own take on *p*. This is a purely positive claim. In asserting it I do not intend to argue that states other than judgment cannot have the relevant feature, an issue on which I will not here take a stance. Since I argue that the type of mineness I am describing is a constitutive feature of the phenomenology of judgment, I will for now refer to it as *judgment-mineness*.

*Judgment-mineness:*

*When engaged in the occurrent conscious act of judging that p, a subject is implicitly and pre-reflectively aware that the judgment is her own, in the sense that it is an exercise of her own agential capacity to make an assessment of p. In virtue of this, she understands the thought as an expression of her epistemic commitment to p.*

There are two key ideas to be picked out of the above description. First, if I am the subject, I am aware when judging that the relevant judgment is *mine* in the sense that it is an exercise of my agency. Second, I am aware that, in being a product of my own agency, the judgment entails my *commitment* to the truth of its content.

I have already defended the first key idea that the phenomenology of judgment necessarily involves regarding the judgment as an exercise of one’s own agency (see 4.4.2). This, I argued, was required for a judgment to entail commitment to a propositional content, in the sense that the content is regarded as true by the subject. This consequently allows the judgment to play the kind of epistemic role we ordinarily expect. Specifically, the subject’s commitment to the content of her judgment allows this judgment to be sensitive to the subject’s wider network of attitudes,

---

<sup>90</sup> I raise this not with the intention of endorsing such a view, but rather to demonstrate one way in which mineness might help shed light on other potential features of conscious experience. For early discussions of this phenomena see Wittgenstein (1958), Shoemaker (1968), and Evans (1982). IEM is discussed in relation to awareness of actions by O’Brien (2007: ch. 11, 2012), and Peacocke (2008: ch. 7).

and to serve as belief formation. However, I now want to defend the idea that, in judging, I am aware not only that *the judgment is an exercise of my agency*, but also that *the judgment that  $p$  entails my commitment to the truth of  $p$* .

As stated, awareness that my judgment is an exercise of my agency is required for me to be epistemically committed to the content of that judgment. However, awareness that my judgment is an exercise of my agency also entails that I *recognise* myself as committed to its content. That is, being aware that my judgment expresses my assessment of a subject-matter entails being aware that I am committed to this assessment. In other words, insofar as I am aware of the judgment that  $p$  as expressing my answer to the question of whether  $p$ , or my take on  $p$ , I understand it as expressing my commitment to  $p$ .

Another way of expressing the thought is this: my being aware that my judgment that  $p$  expresses my agential assessment of  $p$  constitutes my being aware that I am committed to  $p$  (in that I take  $p$  to be true). I cannot regard my judgment as expressing my agential assessment of  $p$  without also understanding it as expressing my commitment to  $p$ .

### 5.3 Judgment-mineness as constitutively grounding self-knowledge of judgments

According to the view I have advanced, a certain kind of phenomenal self-awareness is a *constitutive* feature of judgment in rational human agents. Whenever such a subject judges, an awareness of the fact that the judgment is *hers* is part of what it is for her to make it. Her awareness is a *feature* of her attitude, and not an accompanying attitude. As such, she is not required to take any additional steps to become self-aware. Thus, the awareness is *pre-reflective*. Such an awareness is also *implicit*. By this I mean that a subject is not engaged in explicitly attending to the first-personal nature of her own judgments.

It is my view that the phenomenal awareness we have of our own judgments provides immediate, *constitutive* grounding for knowledge of these judgments. That is, whenever I judge that  $p$ , this judgment is partly constituted by awareness that this judgment is mine, which itself partly constitutes the knowledgeable belief that *I judge that  $p$* . This belief can be thought of as tacit, in that one comes to have it without consciously representing the content that *I judge that  $p$* . Rather, it simply follows from the fact that one judges that  $p$  that one tacitly knows that *one judges that  $p$* .

The notion of tacit belief can be spelled out more clearly.<sup>91</sup> The sort of tacit belief I am concerned with can be distinguished by three key features. However, before I present these, I want to take a moment to highlight one way in which we might refer to knowledge or beliefs as tacit that is distinct from the way in which I am using the term.

One might construe tacit knowledge as knowledge a subject has but is unable to express. For instance, a subject might neither verbally nor in thought be able to express some practical knowledge she has ('knowledge-how', as opposed to 'knowledge-that').<sup>92</sup> A standard example of such a case would be a subject who knows how to stay upright on a bike without being able to express how she does it to herself or others. We might say that this subject has tacit knowledge of how to stay upright on a bike. However, this is *not* the kind of tacit knowledge I am interested in. I now detail the three features which capture the notion of tacit knowledge I am concerned with. Since I understand this knowledge to consist in belief that is true and warranted in a certain way, I will present conditions for tacit belief as opposed to tacit knowledge specifically.

First, I take tacit belief to be *implicit*. I above construed implicitness, in regard to conscious episodes, as a matter of attention. However, the distinction means something slightly different in the case of belief. Having an *explicit* belief, I take it, means having a belief that is consciously manifest in the form of a judgment, or is formed through such a judgment. That is, explicitly believing that *p* involves having a standing commitment to *p* that is occurrently represented in thought. Specifically, in a thought with the content that *p*. Though I have focused on the idea that judgments are normally events of belief formation, judgments also serve the role of making explicit one's standing commitments. The important thing is that judgments, whether they explicitly manifest one's existing beliefs or constitute events of belief formation (or revision), always entail that the subject regards the content of her judgments as true.<sup>93</sup> Thus an implicit belief is just a commitment which is not consciously entertained.

---

<sup>91</sup> The kind of tacit knowledge relevant to my account consists in belief grounded such that it counts as knowledge. This being the case, I will initially present an account of tacit belief in general, before showing why the tacit belief involved in judging counts as knowledge.

<sup>92</sup> This widely accepted distinction was famously made by Ryle (1949).

<sup>93</sup> One might worry that judgments which express beliefs are not actions in the way that judgments which form beliefs are, given that they don't seem to involve active rational assessment. In response to this worry I note that believing that *p* doesn't guarantee that one will judge that *p*, rather it merely disposes one to consciously judging that *p*. It is only the case that a judgment explicitly manifests a belief when the subject's occurrent assessment of *p* matches up with her existing/previously held commitment to *p*. It may also be possible that some conscious thoughts which represent the contents of beliefs are *not* to be understood as judgments, if they do not involve active assessment on the part of the subject. Such thoughts might be better understood as something akin to passing thoughts which are nonetheless connected to one's mental life.

Second, tacit belief seems to consist partly in having the disposition to explicitly think something is the case if prompted in the right kind of way. In other words, if a subject has the implicit belief that  $p$ , the explicit, conscious thought that  $p$  could be induced with appropriate prompts, such as questioning. This kind of thought will involve a particular form of explicit conceptualisation. For instance if the subject who implicitly believes it is evening is asked what time of day it is, she will explicitly deploy her concepts in thinking to herself 'it is evening'.

This would suggest that tacit belief requires having a certain *conceptual capacity*. That is, the subject who believes something tacitly, according to my use of the term, must be in a position to explicitly conceptualise the content of her belief. This requires that she at least *possess* the concepts that would be deployed were she appropriately prompted to consciously think the thing she believes. Returning to our example, the subject would need to possess the concept of 'evening' in order to have the disposition to think 'it is evening', and since such a disposition is a feature of tacit belief, she would need to possess this concept in order to count as having a tacit belief about the time of day.

Third, tacit knowledge involves having a disposition to not only think, but *behave* in a way that manifests that particular knowledge. Consider once again the subject who is implicitly aware of it being evening. Let's suppose that she turns on the television to watch the evening news. Here, her action clearly manifests her knowledge of the time of day. As we saw in 2.2.4, tacit belief plays an *explanatory* role - it can serve to explain the behaviour of a subject.

Given the above features, it seems that a typical case in which a subject has tacit belief that  $p$  would meet the following criteria:

1. The subject has a belief that  $p$ , the content of which she does not consciously think.
2. The subject is disposed to consciously think that  $p$  if prompted in the right way.
3. The subject is disposed to behave in a way that manifests her belief that  $p$ .

I have claimed that judgment-mineness (i.e. the phenomenal self-awareness constitutive of judging) constitutively grounds the tacitly held belief that *I judge that  $p$* . This means that whenever I judge that  $p$ , it is also the case that I believe that *I judge that  $p$* . Having established the three key criteria for tacit belief, I now demonstrate why we should accept that a judging subject is in possession of the belief *that she judges that  $p$* .

To show how the first criterion is met, we must first establish precisely what it means to have a belief. It is clear that *if* the judging subject has the belief that she is judging, it is *not* a belief that is being explicitly manifest in her thought, so the relevant question becomes this: how can we tell if she has the belief? Answering this will require us to draw on the second and third criteria. I have

established that believing consists in being in a standing, non-occurrent, state of commitment to a proposition. That is, to believe that *p*, one must be in a non-occurrent condition of regarding *p* as true. One way to determine if a subject is in such a condition, is by considering whether the second and third criteria are met. Essentially, these criteria help us to determine whether a belief exists, and thus allow us to determine whether the first criterion is met.

I take it that the second criterion is met when one judges. A subject who judges that *p* is disposed to consciously think that *she judges that p* if prompted in the right way. Or at least, she will be in a position to consciously recall her mental act of judging that *p* in such a way that conceptualises the act as her own judgment. In other words, when I judge that *p* I am ordinarily disposed to think that this is what I am doing, or have done, when I am prompted to reflect.

The third criterion is also met. In a straightforward sense, my judging that *p* disposes me to behave in a way that manifests my belief that *I judge that p*, in that it disposes me to *report* this as my act. But a subject will arguably also manifest her belief about her own judgement by engaging in certain forms of critical reasoning. For example, by reflecting critically on evidence relevant to the content of her judgment. The following passage from Shoemaker helps capture the why such behaviour (or at least the disposition to engage in such behaviour) requires self-knowledge of the relevant attitude (1994: 285-286, italics mine):

One does not test propositions at random, or evaluate arguments for arbitrarily chosen propositions. Where it makes sense to engage in such activities is where, for example, the outcome of the investigation will decide which of several ways of rendering one's belief system consistent is the best way; and to know this one must know a good deal about what the contents of that system are [...] What rationalizes the investigation *are one's higher order beliefs about what one believes and has reason to believe.*

Shoemaker's point here is that we engage in critical thinking about certain propositions rather than others because they have some relevance to our epistemic life, and we can only be aware of this relevance by having knowledge of our own commitments. This includes standing epistemic commitments such as beliefs, as Shoemaker notes in the passage, but also occurrent epistemic commitments such as judgments.

I briefly note that cases of tacit belief should be carefully distinguished from those in which a subject is disposed to think something is true without actually believing it to be true. For instance, imagine a subject so absorbed in her work that she fails to register the time of day at all – she does not know whether it is afternoon or evening. If prompted to think about the time of day, she will reflect upon the world and judge (thus forming the belief) that *it is evening*. Here, her

judgment that *it is evening* does not express a belief she had prior to making it. But given the evidence available to her and her general ability to tell what time of day it is based on this evidence, we might think that there is some sense in which she is disposed to have the thought that it is evening.

What the above case demonstrates is that we may encounter situations in which it is difficult to tell whether a subject has tacit belief prior to being prompted in the way described, or whether her thoughts represent newly formed beliefs.<sup>94</sup> But this practical worry about whether or not we can always distinguish cases of tacit belief does not itself serve as an objection to the ontological distinction I have sketched.

Furthermore, it seems that we can often distinguish manifestations of tacit knowledge from judgments which form new beliefs by the *immediacy* with which subjects tend to produce them. To have a tacit belief that *p* is, standardly, to be in a position to immediately think this when prompted. So we can make the second feature of tacit self-knowledge more specific: 'The subject is disposed to *immediately* thinking that *p* if prompted in the right way.' Cases in which anomalous features of a subject's mental life prevent this immediate expression of her belief do not present a challenge to this conception, since they are explained by the very fact that there are extenuating factors at play.

We are now in a position to answer the question at hand, whether a judging subject believes that *she judges*. We have established that a subject who judges that *p* is disposed to immediately consciously think that this what she is doing when prompted. She is also disposed to behave in ways that manifest a standing commitment to the proposition *I judge that p*. So, the second and third criteria are met. Because these criteria are met, we are in a position to attribute the belief that *I judge that p* to the subject who judges that *p*. This, combined with the fact that the judging subject is not explicitly thinking *I judge that p*, means the first criterion (the subject has a belief that *p*, the content of which she does not consciously think) is also met. The significance of this is that we can regard the belief that *I judge that p* as being entailed by judgment that *p*, insofar as it is constitutively grounded in the phenomenal feature of judgment-mineness.

Importantly, it follows that whenever my belief that *I judge that p* is constitutively grounded by my judgment that *p*, this belief will count as knowledge. If my belief that *I judge that p* is grounded by my phenomenal awareness of my judgment, it cannot be false. To be clear, the thought I have presented here is *not* that judgment-mineness is itself knowledge of one's

---

<sup>94</sup> There are likely to be times when this is equally difficult from within the first-person perspective.



judgment. Rather, judgment-mineness is an occurrent form of phenomenal self-awareness which *partly constitutes* my non-occurrent belief that *I judge that p*.

This is not to say that I can *never* falsely believe that I judge that *p*. It is simply that mistakes of this kind can only arise when my belief about my judgment is grounded in something *other* than my phenomenal awareness of my judgment. For instance I might have a false belief that *I judge that p* if I acquire this belief by misremembering my own assessment of *p*. However, in judgment, the constitution relation between judging that *p* and belief that *I judge that p* closes the gap between the two in which error might occur.

The view I have presented here bears a number of similarities with O'Brien's 'agent's awareness' account (2007). According to O'Brien, agent's awareness is a form awareness a subject has of her own actions, including mental actions, as a result of these actions being under her rational control. Or, in other words, it is a type of awareness she has of events that are the product of her rational assessment. This awareness is not derived from observing her own actions or from judging that she is acting. Agent's awareness is a feature of mental action by virtue of this action being a conscious, active assessment.

The core suggestion is that the very idea of an action produced by an active assessment by an agent, carries with it the idea of an assessment by an agent of actions *for her*. For a subject to engage in an assessment of what to do is for a subject to determine what *she* should do. The suggestion then is that any action produced directly on the basis of an active assessment by an agent will be an action of which the agent is aware of as hers. Suppose it is right to say that when a subject judges, doubts or questions one thing rather than another, as a direct result of assessing what to think, she thereby has an awareness of what she is thinking. It is then plausible to suppose that, were the subject to have the appropriate concepts, this would be sufficient to entitle her immediately to self-ascribe her thoughts. (2007: 117)

Given that acting consciously will, in the absence of **any** repressive mechanisms, be sufficient to ground a knowledgeable self-ascription, an agent who acts, and who asks herself what she is doing, will know what she is doing. (2007: 190)

O'Brien takes conscious mental thoughts which are a product of the subject's rational assessment to give the subject immediate warrant to self-ascribe her thoughts. That is, such awareness alone provides warrant for knowledge of one's own mental actions, without the need for 'mediating acts or representations' (2007: 183-4). So one would not, for instance, need to make any kind of inference in order to go from having agent's awareness of one's act to self-attributing the act.

My account bears similarities to O'Brien's, in that it asserts that self-awareness is a feature of the act of judging, and that this grounds knowledge of one's judgment. It is also true on my view that judging involves being aware of what one is doing in such a way that immediately, non-inferentially, grounds knowledge of that judgment. However, on the agent's awareness account, an agent who is  $\phi$ ing and *asks herself* what she is doing knows that she is  $\phi$ ing. On my account it is *not* required that the subject ask herself 'what am I doing?' in order to self-attribute the act of judging.

On my account, a subject's awareness grounds her knowledge of her own judgment in the sense that it partly *constitutes* knowledge that she is judging. That is, on my account, when a subject judges that  $p$ , her phenomenal awareness that *she judges that  $p$*  partially constitutes her belief that *she judges that  $p$* . In other words, the judging subject is not merely warranted, or entitled, to self-ascribe the judgment that  $p$ , she *does* self-ascribe it.<sup>95</sup>

### 5.4 The Epistemology of Transparent Doxastic Self-Knowledge

So far in this chapter I have presented an account of judgment on which a subject who judges that  $p$  knows that *she judges that  $p$* . Such self-knowledge was taken to be grounded by one's phenomenal awareness of one's own judgment, which I identified as *judgment-mineness*. I understood this phenomenal feature in the following way: when engaged in the occurrent conscious act of judging that  $p$ , a subject is implicitly and pre-reflectively aware that the judgment is her own, in the sense that it is an exercise of her own agential capacity to make an assessment of  $p$ . In virtue of this, she understands the thought as an expression of her epistemic commitment to  $p$ .

The account of judgment I have proposed lays the groundwork for my account of transparent doxastic self-knowledge. In this section I will argue that my account of judgment allows us to make sense of transparency, in that it provides an explanation for why world-directed judgments warrant us in self-ascribing beliefs. I will then propose an account of how we go about making transparent self-ascriptions on which transparent self-knowledge is acquired by inference from the premise that *I judge that  $p$*  (which is made available to me purely by judging that  $p$ ) to the conclusion that *I believe that  $p$* . I will argue that such an inference rational, and moreover conducive to knowledge.

---

<sup>95</sup> Another way in which my view importantly differs from O'Brien's is in asserting that one cannot falsely believe that *one judges that  $p$*  when this belief is grounded in phenomenal awareness of one's judgment. O'Brien's account preserves the possibility of falsely self-attributing actions even when one has agent's awareness of them.

#### 5.4.1 A New Inferential Model of Transparent Self-Ascription

I contend that one goes from merely having a transparently derived entitlement to doxastic self-knowledge to transparently self-ascribing belief by way of an *inference*. On my view, a subject acquires doxastic self-knowledge transparently by inferring from the fact that *she judges that p* to the conclusion that *she believes that p*. Her knowledge that *she judges p* is tacitly entailed by her judgment that *p*. For this reason it appears as though the subject is able to infer from *p* to I believe that *p*. It appears this way from the subject's own perspective. Phenomenologically, she transitions from *p* to *I believe that p*. But epistemologically she transitions from the premise *I judge that p*. In this sense, the judgment that *p* acts as a vehicle for the premise *I judge that p*.<sup>96</sup>

On my view, *entitlement* to doxastic self-knowledge comes purely from judging that *p*. One need not actually make the relevant kind of inference in order to have a *right* to transparent doxastic self-knowledge. In this sense, doxastic self-knowledge is not immediate, but knowledge of one's judgments *and* entitlement to doxastic self-knowledge are immediate, both in the sense that they don't rely on inference, and in that they occur at the moment of judgment. Entitlement to doxastic self-knowledge is entailed by knowing that *I judge that p*, which is itself a constitutive feature of judging that *p*.

My view can be clarified by stating more explicitly the difference between having a right to transparent self-knowledge, and using a special transparency method to acquire self-knowledge. Recall that I earlier drew a distinction between the notion of a transparency *procedure*, and transparent *warrant*. On my account, the procedure for learning of one's own beliefs transparently involves inferring from one's judgment that *p* to I believe that *p*. But one is warranted, or entitled, to make this inference simply by judging that *p*. This means that a subject can come to be in a position to make a knowledgeable self-attribution without it being the case that she is trying to acquire self-knowledge; whenever one judges that *p*, one is entitled to self-attribute the belief that *p*. This is something which is left ambiguous by Evans, and is outright rejected by Moran, who holds that a subject must conceive of herself as going about the task of finding out what her belief is in order to acquire doxastic self-knowledge transparently.

So, according to my account there are two relevant kinds of self-knowledge:

---

<sup>96</sup> This rests on the assumption that the phenomenal character of inference need not mirror its epistemic structure. In 5.4.5 I say more to motivate this idea.

**(1) Knowledge that *I judge that p*:**

This *non-inferential* knowledge is tacitly entailed by judgment. Whenever a subject judges that *p*, she knows that *she judges that p*. This is because judgments express the subject's take on *p*, and as such are actions which the subject performs knowingly.

**(2) Knowledge that *I believe that p* (doxastic self-knowledge):**

Doxastic self-knowledge is *inferentially* grounded. It involves inference from the premise 'that I judge *that p*' (which is implicitly contained within the judgment *that p*) to the conclusion 'that I believe *that p*'.

On my view, knowledge that I judge that *p* (1) grounds knowledge that I believe that *p* (2).

Transparent doxastic self-knowledge is acquired by inferring according to the following model:

I judge that *p*

---

I believe that *p*

One might, understandably, question how the model of inference I propose captures a *transparency* method for acquiring doxastic self-knowledge, given that the premise (the basis from which I infer my belief) is a fact about my own mind rather than the subject-matter of my belief. In response I emphasise that one of the most important aspects of my account is that the premise of the inference I describe is made *available* to a subject when she judges that *p*. So, on my account, one's entitlement to transparent self-knowledge comes from making a judgment about the subject matter of her belief.

One might be tempted to draw a parallel between my account and the neo-Rylean account presented by Cassam (2014), according to which one can infer one's own beliefs from one's judgments.<sup>97</sup> However, in spite of superficial similarities, my own view should be regarded as fundamentally distinct from this form of inferentialism. Cassam's view presents a picture of self-knowledge according to which it is not substantively distinct from other forms of knowledge. On my account, doxastic self-knowledge is grounded by a distinctively first-personal form of self-awareness. In this way, transparent doxastic self-knowledge is unlike knowledge of, say, other people's beliefs, the warrant for which we have no special access to. It is crucial on my view that

---

<sup>97</sup> As well as other attitudes and mental phenomena.

we have a special kind of uniquely first-personal *entitlement* to doxastic self-knowledge. This is not so on Cassam's account.

This being said, I recognise it there is a sense in which the account I defend construes doxastic self-knowledge as closer to other knowledge than the transparency accounts I discussed in the first part of the thesis. While self-knowledge of judgments is, on my account, both metaphysically and epistemically unlike other, non-first-personal forms of knowledge, self-knowledge of beliefs is not quite so special. Another subject who had knowledge of my judgment that *p* could very well make the same inference to acquire knowledge of my belief that *p*.

However, it is important that other subjects do not have the same kind of access to my judgments as I do. As a result, other subjects do not earn an entitlement to doxastic self-knowledge at the moment of my judgment. Moreover, I need not be *explicitly* aware of my judgment that *p* as being mine in the relevant sense. From within the first-person perspective I can make inferences from *p* to I believe that *p*, provided that I implicitly understand *p* as something I am epistemically committed to.

#### 5.4.2 How the Alternative View Responds to Boyle's Dilemma

The above model of inference has many benefits over the doxastic schema (3.4.2). In particular, it avoids the major problem presented in Boyle's dilemma. This was the objection that the inference described as the source of transparent self-knowledge is either mad, or else redundant. The inference is mad if it simply involves a transition from a fact about the world to a conclusion about one's attitudes, since facts about the world are not evidence of one's attitudes. In order for the inference not to be mad it would need to be the case that the subject making it already conceived of herself as being epistemically committed to the premise. But in this case the subject seemingly already has doxastic self-knowledge, and so it is unclear why the inference is important – certainly it cannot be understood as grounding doxastic self-knowledge, as such knowledge would exist prior to the inference being made.

But my view does not face Boyle's dilemma. It avoids the 'madness' problem because the implicit premise from which one infers (I judge that *p*) *does* support the conclusion *I believe that p*. As I have shown, to judge that *p* is, *ceteris paribus*, token identical with the event of forming the belief that *p*. Therefore the fact that I judge that *p* provides a sound evidential basis for self-ascribing the belief that *p*. For example, suppose I have knowledge that *I judge that it is snowing* (setting aside for now considerations about how I came to have this knowledge). In knowing this I am aware of myself as consciously affirming the fact that it is snowing. In other words, I know that I

regard it to be true that *it is snowing*. This fact clearly provides good grounds for inferring that *I believe it is snowing*, in a way that the mere fact that it is snowing does not.

*Prima facie*, it may be less clear that my view avoids the redundancy problem. One might think, given that I have described judgment as the event of belief formation, that to know that you judge that *p* just is to know that you believe that *p*. However, knowing that *I judge that p* is not equivalent to knowing that *I believe that p*; knowledge of a mental event normally token identical with forming the belief that *p* is distinct from knowledge of the standing belief that *p*.

The difference between knowing that *I judge that p* and knowing that *I believe that p* is something like the difference between knowing that *it has rained* and knowing that *the street is wet*. Rain is one way for the street to become wet. It is not the only way for this to happen, and under exceptional circumstances the street might remain dry when it rains (e.g. if someone were to cover the pavement with a protective barrier). But, *ceteris paribus* the fact that it has rained provides good evidential grounds for believing that the street is wet. Similarly, judgment is not the only way to form a belief, and in non-standard cases judgment may not result in belief (e.g. if one is under the influence of a motivating bias). However, judgment that *p* normally results in belief that *p*, and so just as knowing that *it has rained* provides good grounds for believing that *the street is wet*, knowing that *I judge that p* provides good grounds for believing that *I believe that p*.

Crucially, like the inference from *it has rained* to *the street is wet*, the inference from *I judge that p* to *I believe that p* is not a redundant transition. In neither case is the premise identical with the conclusion, nor is it required that the subject have prior knowledge of the conclusion in order to make the inference rational.

In my development of Boyle's dilemma, I also raised the following related problem for Byrne's inferentialist account: the fact that *p* does not give the subject a *reason* to self-ascribe the belief that *p*. In this way the inference appears irrational. But the fact that *I judge that p* does give me a reason to self-ascribe the belief that *p*. This is one of the key claims of Peacocke's 'reasons' account of doxastic self-knowledge:

Conscious thoughts and occurrent attitudes can [...] give the thinker reasons for action and judgment. They do so also in the special case in which they give the thinker a reason for self-ascribing an attitude to the content which occurs to the thinker, provided our thinker is conceptually equipped to make the self-ascription. (1998: 71)

Peacocke however denies that one infers from the reason giving fact to the self-ascription (1998: 71-2). His reason for doing so seems to stem from a worry that this makes the epistemology of self-ascriptions of occurrent sensations such as pain mysterious, since the pain gives a reason to

self-ascribe the sensation of pain, but the self-ascription is not made inferentially. I think Peacocke is right to suggest that many of our occurrent experiences (both mental and bodily) can be ascribed non-inferentially. Judgment is one example of this. However, the idea that we shouldn't think of a non-occurrent state like belief as ascribed by inference seems misguided. In 5.4.5 I develop some thoughts about why we shouldn't be afraid of an inferential model of self-knowledge.

We can return for now to the question of whether my inference is rational. I previously suggested that the rationality of an inference seems to depend on an internalist condition being met (3.4.2.3). That is, the fact that a premise normatively supports a conclusion is not the only thing that makes an inference rational. It is also required that the subject herself regard the premise as supporting the conclusion. I argued that it is in order for it to be rational to infer in line with the doxastic schema, the subject therefore would need to take the fact that *p* to support the conclusion that *she believes that p*. But for her to do this would be obviously irrational, since the fact that *p* does not entail, or even make it likely, that one believes that *p*. Thus on Byrne's account, there is no way of meeting the internalist condition on rational inference.

We have already established that the premise of the inference, *I judge that p*, does normatively support the conclusion that *I believe that p*. So the external condition on rationality is met. In order for the inference I describe to be *internally* rational, the subject needs to regard the fact that *she judges that p* as normatively supporting the conclusion that *she believes that p*. I take it that a rational subject in possession of a full concept of belief will in fact be in such a position.

Recall that in judging, a subject is aware that her judgment is an exercise of her agency, and thus that she is committed to its content. I do not take knowledge that one is occurrently committed to a content to be identical with knowledge that one has a belief with that content. However, a subject who understands what it is to believe will regard the fact that she occurrently commits to *p* as normatively supporting the conclusion that *she believes that p*. A subject in possession of the full concept of belief will understand believing as a matter of being in a condition of regarding something as true. She will thus understand her own beliefs as her standing position on the truth of a matter.

A subject who understands her own beliefs in the above way will regard her own judgments, which occurrently express her stance on a subject matter, as telling her something about her standing commitments. The mode of representation is different, but the content is still regarded by the subject as something that is true for her. If she knows that *she judges that p*, i.e. occurrently regards *p* as true, and if she understands belief as the standing condition of regarding

$p$  as true, then she can rationally take the fact that *she judges that  $p$*  as grounds for inferring that she believes that  $p$ .

### 5.4.3 Meeting Conditions for Knowledge

In this section I spell out explicitly how self-ascriptions of belief, made in the way I have proposed, meet plausible conditions for knowledge. I have already suggested that the inference I describe is a rational one. This is a condition that must be met in order for the inference to be knowledge conducive. I will now attempt to give a fuller account of what makes beliefs about my beliefs, acquired by inferring from the fact that *I judge that  $p$* , warranted such that they count as knowledge. I will do so by demonstrating how the following plausible conditions on knowledge are met on my account: normalcy, reliability, and safety.

I first appeal to Martin Smith's 'normic support' account of epistemic justification (2016). According to the account belief are justified by 'normalcy'. Something is 'normal' on this view if we are not inclined to seek further explanation for it. For example, it is normal for the road to be wet after it rains – the road being wet after rain does not require further explanation. If it were to rain and the road remained dry, this would require some further explanation. So, if it rains, I am justified in believing that the road will get wet, since this is what is normically supported by the fact that it is raining.

The normic support conception of justification can be applied to my conception of the judgment-belief relation in the following way. It is normal for judgment to be token identical with belief formation. In other words, a subject's judging *that  $p$*  normically supports her forming the belief *that  $p$* . If the subject were to judge *that  $p$*  and not form the belief *that  $p$* , this would require further explanation. Therefore, if a subject knows herself to have judged *that  $p$* , this justifies her in believing that she has the belief *that  $p$* .

It is important to note that the picture of normic support I have presented is *not* a reliabilist one. To reiterate, the key claim is that the event of a judgment producing a belief *does not require additional explanatory work*. However, the relationship between judgment and belief means that a reliability condition for knowledge can *also* be met when I come to have a belief about some doxastic state of mine on the grounds of knowing my first order judgment about its subject-matter.

To believe that *I believe that  $p$*  on the grounds that *I judge that  $p$*  is reliably knowledge conducive precisely because it is a matter of empirical fact that judgments which do not result in belief (or



express an existing belief) are in the minority. Thus, one can reliably acquire true beliefs about what one's own doxastic states are by inferring from judgment.

Furthermore, to reason from *I judge that p* to *I believe that p* can meet the following modal safety condition: In all nearby worlds where I believe that *I believe that p* on the basis of the fact that *I judge that p*, it is not false that *I believe that p*. Safety only fails in cases where I could easily have failed to form a belief by judging. For instance, this might be the case when I am vulnerable to self-deception. Consider the following example:

I have a desire that my friend be loyal to me, which motivates me to treat evidence of the friend's disloyalty in a motivationally biased way (e.g. by being quick to dismiss such evidence as poor). Now suppose that on one particular occasion I am confronted with evidence strongly suggesting the friend has gone behind my back on some issue. On this occasion I reluctantly judge that my friend has been disloyal, and in doing so form the belief they have been disloyal. It seems that in this case I could easily have failed to form the belief that my friend has been disloyal as a result of my motivating desire that she be loyal. There was therefore a nearby world in which my judging that she is disloyal did not result in the belief that she is disloyal.

I take it that cases such as the one above are rare enough that we need not regard them as posing a threat to the claim that inference from *I judge that p* to *I believe that p* produces beliefs that count as knowledge. However, perhaps it is worth considering whether one could argue that instances of judgment *not* resulting in belief are too common in human agents for the above conditions to be met.

First, does the fallibility of human agents pose a threat to the normalcy condition? If we frequently make mistakes about ourselves due to the influence of biases, or fail to store the contents of our judgments, it might seem problematic to say that judgment is, *ceteris paribus*, token-identical with belief formation.

But epistemic failings, no matter how frequent, would not pose a problem for the normic support conception of justification. As I've established, such mistakes, even if they happen often, would require additional explanation.

'In this sense of 'normal' it could be true that Tim is normally home by six, even if this occurrence is not particularly frequent. What *is* required is that exceptions to this generalisation are always explicable *as* exceptions by the citation of independent, interfering factors—his car broke down, he had a late meeting, he had to detour around roadworks etc.' (Smith, 2016: 39-40)

## Phenomenal Awareness of Judgments

Because all epistemic mistakes, or failings, require explanation, the fallibility of human agents does not pose a problem for the normic support condition.

Next we can consider reliability. Reliability depends upon the process by which one comes to believe producing consistently accurate results. In this way, the frequency with which one forms true beliefs, when employing a given method of belief acquisition, is important. If cases of judgment failing to result in belief were common enough to render the inference from *I judge that p* to *I believe that p* unreliable, then beliefs formed in this way would of course fail to meet a reliability condition on knowledge.

Yet there seems little reason to accept that judgment is unreliable, as a belief-forming method, to the extent that cases of judgment resulting in belief are in the minority. Indeed this would seem to fly in the face of all of our commonplace assumptions about how we come to have standing epistemic commitments. If judgments are not most frequently events of forming or affirming beliefs, then this raises the question of how we can regard ourselves as rational thinkers at all. As Moran emphasises in his account (2001, 2003), part of what it means to be a rational agent is to have beliefs that conform to one's conscious assessments. To deny that these assessments reliably inform and produce standing attitudes would be to deny the rationality of human agents. It would also entail that we cannot, rationally, trust that other agents are in a position to accurately report their own beliefs.

Moreover, we can observe that judgments ordinarily result in standing dispositions to think and behave in ways that manifest one's commitment to the content of those judgments. Accepting an account of belief on which dispositions importantly inform (if not constitute) what it is to believe, we have good reason to regard judgment as the event of belief formation.

Finally, does the possibility of epistemic failure pose a problem for safety? In order for this to be the case, there would need to be the nearby possibility of failing to form the belief that *p* in cases where one's judging that *p* is token identical with belief formation. If this were true, inferring from the premise *I judge that p* to *I believe that p* would not be conducive to safe beliefs.

As a counter to the above worry, we can appeal to the same sorts of considerations I appealed to in defence of reliability. However, while preserving reliability was a matter of showing that the majority of judgments do in fact result in belief formation, holding onto safety requires it to be true that in cases where one's judgment that *p* produces the belief that *p*, one was not *at risk* of going wrong.

If we accept that, in the majority of cases where one forms the belief that *p* by judging that *p*, one could easily have failed to form the belief that *p*, we are forced to accept unattractive

consequences. For instance, it would seem to suggest that there is some element of luck involved in a subject's forming that belief that  $p$ . In undermining the level of control we have over what we believe, which consists largely in our beliefs being determined by our active, conscious assessments, we undermine the level of accountability subjects have for their own beliefs. It would also raise the question of how we could trust such assessments to produce beliefs, if it were the case that judgment could easily fail to determine our standing commitments.

Once again, as was the case for reliability, what's at stake is our ability to justifiably regard ourselves as rational agents. If forming beliefs through judgment were a high risk act, liable to resulting in failure, then we could no longer regard human subjects as capable of authoritatively determining their own standing commitments. We thus have good reason not to regard the possibility of epistemic failure as a threat to the safety of doxastic self-ascriptions grounded in judgment.

#### 5.4.4 Limitations of the Alternative Account

On my view, it is true that if one has an existing belief that  $p$ , but then makes an occurrent judgment that  $\sim p$ , she has grounds for self-attributing the belief that  $\sim p$ . This means that making an occurrent judgment may result in the following cases:

- (1) The subject has an existing belief that  $p$ . She then makes a judgment that  $\sim p$ , which results in her losing the belief that  $p$  and acquiring the belief that  $\sim p$ . She is now entitled to self-ascribe the belief that  $\sim p$ . However, if she were to have made the occurrent judgment that  $\sim p$  with the aim of discovering her belief at the time prior to the judgment, she will be warranted in self-ascribing a different belief from the one she had at the time she decided to figure out what she believed.
- (2) The subject has an existing belief that  $p$ . She makes an occurrent judgment that  $\sim p$ , but because her belief that  $p$  is a recalcitrant, affectively motivated, attitude, she retains the belief that  $p$ . Nonetheless, she is entitled to self-attribute a different belief from the one she in fact possesses.

I take these two cases to demonstrate natural limitations for my transparency method. Basing one's self-attribution on a conscious event of judgment will mean that one is warranted in self-ascribing belief. However, in cases like the first, where a subject judges with the intention of learning her own belief, she will end up learning of the belief she has at that time, rather than the belief she had prior to judgment, when she made up her mind to find out her attitude. So, the transparency method is only conducive to knowledge of beliefs held post-judgment (i.e. presently held-beliefs).

I am happy to accept the above result. When I go about the task of figuring out what I believe, I am concerned simply with learning what my stance on a matter is. The transparency method I describe is conducive to knowledgeable beliefs about one's present doxastic states. If I am interested in finding out what I believed at a specific point in time, there are other ways of securing this knowledge. For instance, by reflecting on the way I behaved and the kinds of thoughts I had at the specified time.

One might wonder, however, why it was a problem for Moran's account that the transparency method could not produce knowledge of some existing beliefs (i.e. some beliefs held prior to deliberation) (1.4.1), but not a problem for my own view. Recall that on my reading of Moran's view, beliefs that meet the rationality condition (in that they conform to one's agent's reasons) must be beliefs that can be learned about using the transparency method (1.1). If they cannot be learned of in this way, Moran suggests they cannot be thought of as rational in the relevant sense (2001: 107-8). In response, I argued that there are some beliefs that meet Moran's rationality condition, but which nonetheless cannot be learned about purely by reflecting upon one's agent's reasons. This served to undermine the thought captured above, that beliefs that meet the rationality condition must also meet the transparency condition.

However, on my own view, I *do not* commit to the idea that all rational, reason-responsive beliefs can be learned about via the inferential transparency method I describe. Rather, my view simply states that whenever a subject makes a conscious judgment that *p*, she has warrant to self-ascribe the belief that *p*. A subject's conscious assessment of her reasons, when it results in judgment, puts her in a position to know what she believes because judging normally *is* belief formation. But it is not a condition on rational belief that the subject can learn about a belief by inferring from her conscious judgment. A belief's being rational simply makes it likely that, under normal circumstances, a subject will be in a position to learn of the belief in this way. Thus it is not a problem for my view if there are some rational beliefs a subject cannot learn about transparently.

The method I describe will not, as Moran's account does on my reading, fail to produce an answer in cases where the subject has an attitude towards *p* but regards her reasons as supporting two different judgment-conclusions. This is because, on my view, entitlement to self-knowledge explicitly comes from making a judgment on a subject-matter itself, and not from a conclusion about what one's reasons support.

The second case I considered represents a way in which the transparency method I propose can yield an incorrect answer. In other words, it shows one way in which the method is fallible. Since judgment and belief can come apart in certain anomalous cases, beliefs about one's doxastic states grounded in knowledge of one's own judgments will not always count as knowledgeable.

This should not be considered a problem for my view. On the contrary, it is a benefit of my account that it allows for the possibility of mistakes. Since subjects are not perfectly rational, in that their world directed judgments do not always line up with their standing beliefs, it would be implausible to suppose that one can always get the right answer when reasoning transparently.

My account offers one form of explanation for how human agents go wrong when attributing beliefs to themselves. Importantly however, in cases where a subject does come to hold a false belief about her doxastic state, if this belief is based on her knowledge of a judgment she has made with the same contents, the false belief is still epistemically warranted. Peacocke's university administrator (1998: 90), for example, may well have a false belief about her standing doxastic attitude towards degrees gained abroad, but if she has this belief on the basis of her occurrent judgment on the matter, then the belief is warranted, even though it lacks the status of self-knowledge.

#### 5.4.5 Why Not Inference?

It is widely held that if we are to do justice to the distinctive nature of doxastic self-knowledge, we *cannot* think of it as inferentially acquired. As we have seen, Moran (2001, 2003), Boyle (2009, 2011), O'Brien (2007), and Peacocke (1998), whose views have been the central inspiration for my own account, all reject that doxastic self-knowledge is the product of inference. As is clear from the account I have advanced, I do not share their averseness.

While I have already demonstrated that an inferential account of self-knowledge needn't face the problems associated with Byrne's inferentialism, I here give a further statement in defence of understanding transparent doxastic self-knowledge as inferentially acquired. What I say is not intended to be the final word on this topic. Rather, I here have the more modest aim of showing that inference need not be a dirty word when it comes to self-knowledge of beliefs.

We can begin by noting that part of what motivates the widespread rejection of an inferential model of doxastic self-ascription is the thought that such a model fails to properly distinguish self-knowledge from other knowledge, including knowledge of other people's beliefs. The idea, broadly, seems to be that if we need to make an inference to learn of our own beliefs, then we are disconnected from these beliefs in such a way that undermines our authority over them.

I want to suggest that the contemporary aversion to inferential doxastic self-knowledge is partly a result of the way in which inferential accounts are ordinarily advanced. In particular, the more traditional Rylean accounts (see 0.2.2) have done much to reinforce the idea that inferential self-knowledge cannot be distinctive. Byrne's account at least attempts to show that inferential self-

ascriptions are unlike other forms of inference, in that they have the features of peculiar and privileged access (see 3.2). But as we saw, this view is wanting in that the inference it posits seems irrational.

Yet our options in advancing an inferential account of doxastic self-knowledge are not limited to proposing an inference that is apparently irrational, or else renders self-knowledge undistinctive. In my own account I showed that we have a distinctive form of non-inferential self-knowledge of our judgments, and that this special form of tacit knowledge gives us a distinctive kind of entitlement to inferential self-knowledge. Thus doxastic self-knowledge on my account is inferential, but it is warranted in a way that makes it quite unlike other knowledge. Moreover, the inference I describe is a rational one (see 5.4.2). Thus my inferential account avoids the problems associated with both the Rylean views, and Byrne's transparency account.

Furthermore, while it is intuitive to think that we are immediately (i.e. non-inferentially) aware of occurrent mental events, especially when these events are exercises of our agency, it is less intuitive to think that we can have immediate awareness of our non-occurrent mental states. On my understanding, to believe something is to be in a standing condition of regarding  $p$  as true. Beliefs are not, strictly speaking, the kind of thing that one can experience directly. What one experiences are manifestations of belief in one's conscious thoughts, behaviour, and actions. As I have attempted to argue in Part 2 of this thesis, it is for this reason that conscious judgments entitle us to doxastic self-knowledge. Given that we can only have occurrent experiences of the events connected to attitudes like belief, and not of the standing attitudes themselves, it is not unreasonable to suppose that we learn about our standing attitudes by inferring them from such occurrent events.

Importantly, we need not adopt an overly demanding conception of inference. Such reasoning need not take place over an extended period of time. We can conceive of it minimally, as simply a transition from one believed content to another believed content.<sup>98</sup> Inferences can happen both spontaneously (i.e. without the subject deciding to engage in such reasoning) and immediately, in the temporal sense, since one can infer a conclusion immediately, or automatically, upon consciously grasping a premise.

For instance, I might infer from the fact that *my father's car is outside of his house* that *my father is at home*, but it needn't be that this event consists of an event of consciously thinking to myself, *my father's car is outside, if my father's car is outside then my father is home, therefore my father is home*. Rather, it seems that I infer that *my father is home* immediately upon grasping that *his*

---

<sup>98</sup> I assume for now that the subject must regard the premise as evidentially supporting the conclusion.

*car is outside*. To put this in slightly different terms, while inferences might be formally structured as syllogisms with discrete steps, the phenomenology of inference need not mirror their formal structure.

Boghossian's account of theoretical reasoning may be helpful here (2014). On his account, inference is regarded as taking place within the theoretical domain: to infer is to reason with beliefs. This is compatible with my account, on which a subject reasons from the belief that she judges that *p* to the believed conclusion that *she believes that p*. According to Boghossian's view, much of our reasoning, including inferences, falls somewhere in between what Daniel Kahneman identifies as 'System 1' and System 2' reasoning (Kahneman, 2011: 20-21):

System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control.

System 2 allocates attention to the effortful mental activities that demand it, including complex computations. The operations of System 2 are often associated with the subjective experience of agency, choice, and concentration.

Boghossian takes these distinctions to correspond to the distinction between 'reasoning that is sub-personal, sub-conscious, involuntary and automatic, on the one hand, and reasoning that is person-level, conscious, attention hogging and effortful, on the other' (2014: 2).<sup>99</sup> Importantly for our purposes, while he regards inferences such as the one I make about my father as having elements of System 2 thinking, in that they are personal level, conscious, mental actions, he also rightly points out that they have elements of System 1 thinking, in that they are 'quick, relatively automatic and not particularly demanding on the resources of attention' (2014: 2).<sup>100</sup> Boghossian consequently suggests that inferences are instances of what we might call 'System 1.5' reasoning and above.

We can think of the inference involved in making self-ascriptions of belief as having the status of System 1.5 reasoning. It is, simply, a rational transition between contents, the latter of which the subject takes to follow from the first. It does not need to be conceived as a demanding rational process during which a subject is reflectively cognisant of the formal structure of her reasoning. The transition from the premise that *I judge that p* (made available by the judgment that *p*) to the

---

<sup>99</sup> There are clear connections between the distinctions discussed here and my earlier account of mental action. While it may well be worth pursuing the question of how these distinctions bear upon my account of judgment, I will treat this as a topic for discussion at another time.

<sup>100</sup> Boghossian also claims that such inferences are voluntary – I understand this as the claim that they are voluntary *under one description*, i.e. in that I voluntarily engage in the reasoning, and not in that I can will myself to draw conclusions with specific contents.

self-ascription of belief that  $p$  can happen quickly – potentially automatically – without concerted concentration on the premise content.

As a final note, it should be observed that thinking of transparent doxastic self-knowledge as inferential helps make the epistemology of such knowledge less mysterious. One of the key problems for Moran's view was that it failed to clearly capture how one actually makes a self-ascription of belief on the basis of one's reasons assessment. On Boyle's view, the self-ascription was built into the belief itself. Among other problems, this had the unintuitive consequence that we know our own beliefs even when we are apparently ignorant of them (in that we are unable to make a conscious self-ascription). Boyle's account also relied on the somewhat obscure notion that we make explicit self-ascriptions by 'reflecting' on our condition. I.e., by attending to our belief state so as to make our knowledge of it conscious. Yet, as I have suggested here, it is unclear how we can turn our reflection upon a state. Rather, it seems that what we consciously attend to are the occurrent mental events connected to a state.

By understanding transparent self-knowledge as the product of an inference, from a premise which can rationally be taken in support of the conclusion, my account gives a relatively straightforward explanation of the transition from judging that  $p$  (in which one knows that *one judges that  $p$* ) to self-ascribing the belief that  $p$ . In this sense, my account has something in common with the accounts that construe self-knowledge as modelled on other knowledge (see 0.2), since the inference I describe could be used to gain knowledge of other people's beliefs. That is, I could infer from the fact that *my friend judges that  $p$* , that *my friend believes that  $p$* , and this ascription, if true, would count as knowledge. But crucially, I would not be entitled to make this inference just in virtue of my friend judging that  $p$ . For this reason, I cannot have transparent knowledge of others' beliefs.

To summarise the key thoughts in this section, we need not think that accepting an inferential picture of doxastic self-knowledge requires sacrificing either rationality or distinctiveness. We can think of ourselves as having a distinctive form of warrant for doxastic self-knowledge, while also holding that it is inferentially acquired. Moreover, thinking of the knowledge of our own beliefs as inferentially acquired does not undermine the sense in which we are agentially responsible for them. We can still think of beliefs as the product of our rational assessments. My account also allows that we have immediate, distinctively first-personal knowledge of the *event* of belief formation, when this event is token identical with judgment. It has the further benefit of giving a more straightforward account of doxastic self-ascription (i.e. the transparency method) than the agency-based views.



## Conclusion

I began this thesis with the following aim: to find an account of doxastic self-knowledge that could do justice to the distinctive nature of our knowledgeable self-ascriptions, while also giving a satisfying explanation of how these self-ascriptions are warranted. That is, an explanation of what makes them count as knowledge. I sought to accomplish this aim by finding an account that could explain the epistemology of ‘transparent’ self-knowledge, i.e. knowledge grounded in one’s world-directed judgments.

In Part 1 of the thesis I considered three existing accounts that purport to explain the source of our warrant for transparent self-ascriptions of belief. In Chapter 1 I discussed Moran’s highly influential agentalist account (2001, 2003). This account made a number of insightful observations about the relationship between rational agents and beliefs (1.1). In particular, it drew attention to the fact that a subject’s authority over her own beliefs consists not merely in the fact that she is better placed to accurately identify these beliefs than other subjects, but in the fact that she herself plays a role in bringing them about. Specifically, by exercising her rational capacity to assess her reasons and form epistemic commitments through these assessments.

However, while the agentalist view made a number of suggestive and plausible claims about the agent-belief relationship, it failed to give a satisfying account of how reflection upon one’s reasons provides grounding for knowledgeable self-ascriptions. I argued that deliberation, on Moran’s account, must involve understanding one’s reasons as favouring certain epistemic responses (including belief, disbelief, and suspension of belief). However, in order for deliberation to provide warrant for doxastic self-ascriptions, it would need to be the case that a subject could only regard her reasons as permitting a single attitude response. In this way, the account appeared committed to what I identified as a form of agent-relative uniqueness (1.4.1). Given that a subject could regard her beliefs as permitting more than one epistemic response, it seemed that reflection upon these reasons might not always be able to yield knowledge of rational beliefs.

Moreover, Moran’s transcendental explanation for the source of our warrant relied upon an overly demanding account of rational agency (1.4.2). Specifically, the account suggested that subject must not only be responsive to reasons in deliberation, but also be self-aware. Specifically, the account seemed to require the subject to be aware of the following: that her beliefs conform to her reasons (the transcendental assumption), that in deliberation she is engaged in belief-formation, and that deliberation is a means by which she can learn what she believes. I suggested that the account is therefore dependent on an overly reflective form of rational agency to explain what it is that allows our rational reasons assessment to ground our doxastic self-ascriptions.

## Conclusion

Boyle's reflectivist account addressed the gaps in Moran's agentialism. Like Moran's account, reflectivism also understood a subject's right to transparent self-knowledge as deriving from her capacity for rational agency. Specifically, the account stated that a subject has knowledge of her own beliefs in virtue of the fact that beliefs are themselves active exercises of rational agency (2.1). The agentialness of belief was taken to consist in the fact that believing is constituted by endorsing belief that  $p$  – i.e. regarding  $p$  as to be believed. This supposedly made it the case that self-knowledge is *constitutive* of believing. In this way, reflectivism gave an answer to the key question left open by Moran's account, of how one's active rational assessments ground self-ascriptions.

However, reflectivism faced its own problems. The unusual metaphysical conception of belief it advanced bucked widely held ideas about the nature of non-occurrent states. In doing so, it made belief into an overly-reflective condition. The notion of belief being constituted by endorsement also gave rise to the possibility of a vicious regress (2.2.3). Moreover, the account was forced to give problematic accounts of epistemic failings. Self-ignorance in repressed belief, for instance, was explained by the fact that a subject lacked access to her knowledge of her own belief, even though self-ignorant subjects tend not to display dispositions which manifest such knowledge (2.2.4.1).

The inferentialist account advanced by Byrne proposed an interesting alternative to the agency-based views. However this account fared worst of all under scrutiny. The account proposed that transparent self-knowledge is acquired by inferring from a fact about the world,  $p$ , to the conclusion *I believe that  $p$*  (3.1). This inference was taken to be rational and knowledge conducive because making it guaranteed the truth of the premise. However, an assessment of the inference revealed its failure to meet numerous widely accepted conditions on rational inference (3.4.2). I claimed that in light of the problems with this non-agential approach, and the broad credibility of ideas developed in the agency-based approaches, we should pursue an account of transparent self-knowledge which understands this knowledge as grounded in agency. I suggested that we take the key insights of the agency-based accounts as a foundation from which to develop an alternative view.

In Part 2 I outlined an alternative to the existing accounts of transparent self-knowledge. I believe that this view, while still in need of refinement, can provide a way of understanding the source of our warrant for doxastic self-knowledge that avoids the problems associated with the agentialist and reflectivist accounts. This view takes a central idea of Moran's original agency account – that the nature of active, rational reasons assessment is fundamentally self-aware – and develops it by appealing to the nature of *conscious* judgment.

By focusing on the mental event of judgment, I regard my project as offering a more developed account of what is construed as *deliberation* on Moran's agentialist view (though my notion of judgment is a markedly more minimalist conception of *reasons assessment* than deliberation). In regarding a subject's judgments explicitly as *conscious events of belief formation* and *mental actions*, we can give credence to the idea that they are partly constituted by self-awareness. Such a conception also allows us to do justice to our accountability for our doxastic states, without leaning on an overly reflective notion of belief.

A phenomenal approach to the self-awareness characteristic of judgment strikes me as a promising route to take. There is already a wealth of literature concerned with implicit, non-reflective awareness of conscious episodes. As I hope to have demonstrated, the ideas contained within this literature can be utilised to help us understand how the conscious judgment that *p* puts us in a position to make a knowledgeable doxastic self-ascription. Our judgments give warrant for self-ascriptions of non-occurrent belief states because, in judging, I already knowledgeably self-ascribe an occurrent commitment.

One of the most difficult questions my account faces is how exactly phenomenal awareness that *a judgment is mine* grounds my knowledgeable belief that *I judge that p* (which in turn provides warrant for doxastic self-ascription). I have here proposed a constitutive explanation of this relation, though it is one which will need to be drawn out more precisely in any refined versions of the account.

Ultimately, my hope is to have built a convincing case for thinking of one's warrant for transparent self-ascriptions of belief as inextricably linked to one's capacity not just to engage in rational assessment of one's reasons, but to do so with an understanding of these assessments *as one's own*. It is this fact, I contend, that makes it the case that I am in a position to self-ascribe beliefs not just when I ask myself, what do I believe? Rather, I am in a position to knowledgeably self-ascribe beliefs whenever I make a conscious assessment of my reasons. In this way, doxastic self-knowledge is always available to me, provided that I am actively, rationally, engaged with the world.



## References

- Armstrong, D. M., 1993. *A Materialist Theory of the Mind*. Abingdon: Routledge.
- Bar-On, D. & Ochs, J., 2018. The Role of Inner Speech in Self-Knowledge: Against Neo-Rylean Views. *Teorema*, 37(1), pp. 5-22.
- Boghossian, P., 2014. What is Inference?. *Philosophical Studies*, 169(1), pp. 1-18.
- BonJour, L. & Sosa, E., 2003. *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. Malden: Wiley-Blackwell.
- Boyle, M., 2009(a). Active Belief. *Canadian Journal of Philosophy*, Volume 39, pp. 119-147.
- Boyle, M., 2009(b). Two Kinds of Self-Knowledge. *Philosophy and Phenomenological Research*, Volume 78, pp. 133-164.
- Boyle, M., 2011(a). 'Making up Your Mind' and the Activity of Reason. *Philosophers' Imprint*, 11(17), pp. 1-24.
- Boyle, M., 2011(b). Transparent Self-Knowledge. *Proceedings of the Aristotelian Society*, Volume 85 (Supplementary), pp. 223-241.
- Bratman, M., 1987. *Intention, Plans, and Practical Reason*. Cambridge: Harvard University Press.
- Broome, J., 2013. *Rationality Through Reasoning*. Malden: Wiley-Blackwell.
- Byrne, A., 2005. Introspection. *Philosophical Topics*, Volume 33, pp. 79-104.
- Byrne, A., 2011. Transparency, Belief, Intention. *Proceedings of the Aristotelian Society*, Volume 85 (Supplementary), pp. 201-221.
- Byrne, A., 2018. *Transparency and Self-Knowledge*. Oxford: Oxford University Press.
- Campbell, J., 1999. Schizophrenia, the space of reasons, and thinking as a motor process. *The Monist*, 82(4), pp. 609-25.
- Carruthers, P., 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.
- Cassam, Q., 2014. *Self-Knowledge for Humans*. Oxford: Oxford University Press.
- Casullo, A., 2007. What is Entitlement?. *Acta Analytica*, Volume 22, pp. 267-279.

## References

- Chrisman, M., 2018. Epistemic Normativity and Cognitive Agency. *Noûs*, 52(3), pp. 508-529.
- Coliva, A., 2002. Thought insertion and immunity to error through misidentification. *Philosophy, Psychiatry, and Psychology*, 9(1), pp. 27-34.
- Coliva, A., 2008. Peacocke's Self-Knowledge. *Ratio*, 21(1), pp. 12-27.
- Coliva, A., ed., 2012. *The Self and Self-Knowledge*. In: Oxford: Oxford University Press.
- Drake, J., 2017. Doxastic Permissiveness and the Promise of Truth. *Synthese*, 194(12), pp. 4897-4912.
- Frankfurt, H., 1977. Identity and Externality . In: A. Rorty, ed. *The Identities of Persons*. Berkeley : University of California Press, pp. 239-51.
- Gendler, T. S., 2008. Alief and belief. *Journal of Philosophy*, Volume 105, pp. 634-663.
- Gertler, B., 2001. Introspecting Phenomenal States. *Philosophy and Phenomenological Research*, Volume 63, pp. 305-328.
- Gertler, B., 2011. Self-knowledge and the Transparency of Belief. In: A. Hatzimoysis, ed. *Self-Knowledge*. Oxford: Oxford University Press, pp. 125-145.
- Gertler, B., 2018. Self-Knowledge and Rational Agency: A Defense of Empiricism. *Philosophy and Phenomenological Research*, 96(1), pp. 91-109.
- Guillot, M., 2017. I Me Mine: on a Confusion Concerning the Subjective Character of Experience. *Review of Philosophy and Psychology*, 8(1), pp. 23-53.
- Heal, J., 2002. On First-Person Authority. *Proceedings of the Aristotelian Society*, 102(1), pp. 1-19.
- Hieronymi, P., 2006. Controlling Attitudes. *Pacific Philosophical Quarterly*, 87(1), pp. 45-74.
- Hieronymi, P., 2008. Responsibility for Believing. *Synthese*, 161(3), pp. 357-373.
- Hieronymi, P., 2009. Believing at Will. *Belief and Agency*, Volume 35, pp. 149-187.
- Hieronymi, P., 2014. Reflection and Responsibility. *Philosophy and Public Affairs*, 42(1), pp. 3-41.
- Husserl, E., 1963. *Ideas: A General Introduction to Pure Phenomenology (1913)*. New York: Collier Books.
- Kahneman, D., 2011. *Thinking Fast and Slow*. London: Allen Lane.

- Kopec, M. & Titelbaum, M. G., 2016. The Uniqueness Thesis. *Philosophy Compass*, Volume 11, p. 189–200.
- Korsgaard, C. M., 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.
- Kriegel, U., 2007. Consciousness: Phenomenal consciousness, access consciousness, and scientific practice. In: P. Thagard, ed. *Handbook of philosophy of psychology and cognitive science*. Amsterdam: Elsevier, pp. 195-217.
- Kriegel, U., 2009. *Subjective Consciousness*. Oxford: Oxford University Press.
- Levine, J., 2001. *Purple haze: The puzzle of consciousness*. Oxford: Oxford University Press.
- Lycan, W. G., 1996. *Consciousness and Experience*. Cambridge: MIT Press.
- Macdonald, C., 1999. Shoemaker on Self-Knowledge and Inner Sense. *Philosophy and Phenomenological Research*, 59(3), pp. 711-738.
- Martin, M. G. F., 1998. An Eye Directed Outward. In: C. Wright, W. Barry & C. Macdonald, eds. *Knowing our Own Minds*. Oxford: Clarendon Press, pp. 99-121.
- McHugh, C., 2011. Judging as a Non-Voluntary Action. *Philosophical Studies*, 152(2), pp. 245-269.
- McHugh, C., 2013. Epistemic Responsibility and Doxastic Agency. *Philosophical Issues*, 23(1), pp. 132-157.
- McHugh, C. & Way, J., 2018. What is Reasoning?. *Mind*, 127(505).
- Mellor, C. S., 1970. First Rank Symptoms of Schizophrenia. *The British Journal of Psychiatry*, 117(536), pp. 15-23.
- Moran, R., 1997. Self-Knowledge: Discovery, Resolution, and Undoing. *European Journal of Philosophy*, Volume 5, pp. 141-161.
- Moran, R., 2001. *Authority and Estrangement*. Princeton: Princeton University Press.
- Moran, R., 2003. Responses to O'Brien and Shoemaker. *European Journal of Philosophy*, 11(3), pp. 402-419.
- Moran, R., 2005. Getting Told and Being Believed. *Philosophers' Imprint*, 5(5), pp. 1-29.
- Nagel, T., 1974. What Is It Like to Be a Bat?. *The Philosophical Review*, 83(4), p. 435–50.

## References

- O'Brien, L., 2003. Moran on Agency and Self-knowledge. *European Journal of Philosophy*, Volume 11, pp. 375-390.
- O'Brien, L., 2007. *Self-Knowing Agents*. Oxford: Oxford University Press.
- Peacocke, C., 1996. Our Entitlement to Self-Knowledge Part 2. *Proceedings of the Aristotelian Society*, Volume 96, pp. 117-158.
- Peacocke, C., 1998. Conscious Attitudes, Attention and Self-Knowledge. In: C. Wright, W. Barry & C. Macdonald, eds. *Knowing our Own Minds*. Oxford: Clarendon Press, pp. 63-98.
- Peacocke, C., 1999. *Being Known*. Oxford: Oxford University Press.
- Peacocke, C., 2007. Mental Action and Self-Awareness (I). In: J. Cohen & B. P. McLaughlin, eds. *Contemporary Debates in the Philosophy of Mind*. Oxford: Blackwell, pp. 358-376.
- Peacocke, C., 2009. Mental Action and Self-Awareness (II): Epistemology. In: L. O'Brien & M. Soteriou, eds. *Mental Action*. Oxford: Oxford University Press, pp. 192-215.
- Peacocke, C., 2014. *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness*. Oxford: Oxford University Press.
- Pink, T., 2003. Action and self-determination. *Intellectica*, Volume 36–37, p. 247–259.
- Pitt, D., 2004. The Phenomenology of Cognition or "What is it Like to Think that P?". *Philosophy and Phenomenological Research*, 69(1), pp. 1-36.
- Reed, B., 2010. Self-knowledge and Rationality. *Philosophy and Phenomenological Research*, 80(1), pp. 164-81.
- Reed, B., 2010. Self-knowledge and Rationality. *Philosophy and Phenomenological Research*, Volume 80, pp. 164-181.
- Rödl, S., 2007. *Self-Consciousness*. Cambridge: Harvard University Press.
- Roessler, J., 2013. The silence of self-knowledge. *Philosophical Exploration*, 16(1), pp. 1-17.
- Russell, B., 1912. *The Problems of Philosophy*. 2001 ed. Oxford: Oxford University Press.
- Ryle, G., 1949. *The Concept of Mind*. 2000 ed. Wiltshire: Penguin Classics.
- Sartre, J. P., 1957. *La transcendance de l'ego (1936)*. New York: The Noonday Press.



- Scanlon, T. M., 2010. *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge: Harvard University Press.
- Schoenfield, M., 2014. Permission to Believe: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief. *Noûs*, Volume 48, p. 193–218.
- Setiya, K., 2008. Believing at Will. *Midwest Studies in Philosophy*, 2(1), p. 36–52.
- Shah, N. & Velleman, J. D., 2005. Doxastic Deliberation. *Philosophical Review*, 114(4), pp. 497-534.
- Shoemaker, S., 1994. Self-knowledge and 'Inner Sense'. *Philosophy and Phenomenological Research*, Volume 54, pp. 249-314.
- Shoemaker, S., 2003. Moran on Self-Knowledge. *European Journal of Philosophy*, Volume 3, pp. 391-401.
- Shoemaker, S., 2009. Self-Intimation and Second Order Belief. *Erkenntnis*, Volume 71, pp. 35-51.
- Smith, A. M., 2005. Responsibility for Attitudes: Activity and Passivity in Mental Life. *Ethics*, 115(2), pp. 236-271.
- Smith, A. M., 2008. Control, Responsibility, and Moral Assessment. *Philosophical Studies*, 138(3), pp. 367-392.
- Smith, M., 2016. *Between Probability and Certainty*. Oxford: OUP.
- Stephens, G. L. & Graham, G., 2000. *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*. Cambridge: MIT Press.
- Way, J., 2007. Self-Knowledge and the Limits of Transparency. *Analysis*, Volume 67, pp. 223-230.
- Wedgwood, R., 2002. The Aim of Belief. *Nous*, 36(s16), pp. 267-297.
- Wedgwood, R., 2012. Justified Inference. *Synthese*, 189(2), pp. 1-23.
- White, R., 2005. Epistemic Permissiveness. *Philosophical Perspectives*, Volume 19, pp. 445-59.
- White, R., 2013. Evidence Cannot Be Permissive. In: M. Steup, J. Turri & E. Sosa, eds. *Contemporary Debates in Epistemology 2nd ed*. Hoboken: John Wiley & Sons, p. 312–23.
- Williams, B., 1973. Deciding to Believe. In: B. Williams, ed. *Problems of the Self*. Cambridge: Cambridge University Press, p. 136–51.
- Wilson, T. D., 2002. *Strangers to Ourselves*. Cambridge: Belknap Press.

## References

Zahavi, D., 1999. *Self-awareness and alterity: A phenomenological investigation*. Evanston: Northwestern University Press.

Zahavi, D., 2005. *Subjectivity and selfhood: Investigating the first-person perspective*. Cambridge: MIT Press.

Zahavi, D. & Kriegel, U., 2015. For-me-ness: What it is and what it is not. In: D. O. Dahlstrom, A. Elpidorou & W. Hopp, eds. *Philosophy of mind and phenomenology: Conceptual and empirical approaches*. London: Routledge, pp. 36-53.