Knowing Yourself is Something You Do

By

Antonia Michelle Rosen Peacocke

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Philosophy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John Campbell, Co-chair
Professor Barry Stroud, Co-chair
Professor Tania Lombrozo

Spring 2018

Abstract

Knowing Yourself is Something You Do

by

Antonia Michelle Rosen Peacocke

Doctor of Philosophy in Philosophy

University of California, Berkeley

Professors John Campbell and Barry Stroud, Co-Chairs

Why do your self-attributions of beliefs and intentions ordinarily constitute authoritative self-knowledge? You can self-attribute a belief or an intention transparently. For instance, you can transparently self-attribute a belief that *p* by judging that *p*. You can transparently self-attribute an intention to Φ by deciding to Φ. However, recognizing just this much does not completely explain the epistemology of transparent self-attributions.

Self-attributions of this kind count as authoritative knowledge because they involve a form of practical knowledge. You can intentionally control the kind of attitude you take up in conscious thought, and when you do that, you know what kind of attitude you are taking up in conscious thought. Then, in the context of transparent self-attribution of belief or intention, a judgment that *p* or a decision to Φ can have a complex identity. A judgment that *p* can also *be* a self-attribution of a belief that *p*, and a decision to Φ can also *be* a self-attribution of an intention to Φ. To explain how this can be the case I introduce the linked notions of embedded mental action and content plurality.

The view of self-knowledge that emerges also explains why there are contents involving belief attributions that are absurd to assert or to judge even though they can be true. These contents are Moorean absurdities for belief. I argue that there are no corresponding Moorean absurdities for intention, even though you also have transparent self-knowledge of what you intend to do. This points to an important attitudinal distinction between belief and intention: intentions are not beliefs.

The difference between first-personal and third-personal methods of attributing attitudes is subtle. The specialness of the first-personal perspective cannot be explained in terms of epistemic groundlessness, as many have tried to do. You must also make third-personal groundless attributions of belief to understand others' intentional behavior.

Despite philosophical skepticism on this point, transparent self-knowledge really is valuable, in a special sense. Having complete diachronic transparent self-knowledge involves having no hidden attitudes and having a diachronically unified self of the kind that is required for evaluation in terms of authenticity.

The epistemology of self-knowledge relies crucially on the fact that you can *do* things in thought. Knowing yourself is something you do because intentional action is indispensable to authoritative, knowledgeable self-attribution of beliefs and intentions.

Knowing Yourself Is Something You Do

Table of Contents

Acknowledgements

I'm deeply grateful to have had Barry and John as teachers and advisers on this project over the past few years. Barry has helped me develop a philosophical compass that guides me towards real questions. John has shown me how to find the spark of life in a philosophical project. Their influence on my thought can hardly be overstated.

Thanks also to Tania Lombrozo, Geoff Lee, Mike Martin, and Alva Noë for their reading, commentary, discussion, support, analysis, and guidance in this project. Each of them has challenged me in distinct and unforgettable ways.

To my original family, thanks: to Dad, for on-call philosophical advice of the kind everyone needs but not everyone is so lucky to have; to Mom, for grounding and humor and consistent commentary of the unfailingly positive kind; and to Aley, for unflagging patience and kindness. And to my new family—Jack, Anne, Liz, Brady, Sarah, Lee, and the kiddos—thank you for your constant support and genuine interest in my project.

The graduate student community at Berkeley has been immeasurably helpful. I'd like to thank everyone who has been here while I have and who has helped me to grow my work in new directions. I am especially grateful to those who have patiently read and commented on my work, including Jim Hutchinson, Ravit Dotan, Arc Kocurek, Jeff Kaplan, Austin Andrews, Caitlin Dolan, Joe Kassman-Tod, Eugene Chislenko, Adam Bradley, Julian Jonker, Quinn Gibson, Alex Kerr, and Kirsten Pickering.

The commentary I've received from readers outside the department has been fantastically helpful. Many thanks to Susanna Siegel, Ram Neta, Brie Gertler, Katia Samoilova, Arden Koehler, Philippe Chuard, Dimitri Mollo, Edgar Phillips, and Cherry Brice. I'm also very grateful to a thorough and exacting anonymous reviewer at *Philosophical Studies*.

I learned a lot from presenting parts of the material in this dissertation in various venues. I'd like to thank audiences at Stanford University, Harvard University, Queen's University, Ryerson University, University of London, the 2016 Pacific Meeting of the American Philosophical Association, and the 2017 Workshop on Introspection and Self-Knowledge in Mexico City.

I have been lucky to have friends outside my department who pick me up and cheer me on. I don't know what I did to deserve you all: Sarah, Arden, Sara, Andrew, Cherry, Emily, Susan, Jess, Eve, Rachel, Adi, Kate. Thank you for putting up with me.

Most of my gratitude is due to Jackson—although, and because, he refuses to accept it.

Introduction. First-Person Authority and Transparency

This dissertation is about self-knowledge. In particular, it is about how you know what you believe and how you know what you intend to do. It is also about why it can be a matter of deep personal importance to know what you believe and what you intend.

I have chosen to focus on knowledgeable self-attribution of belief and intention rather than other states, acts, or experiences for a few reasons. There are important commonalities between these states. Beliefs and intentions are normatively and descriptively individuated, though in different ways. They are states of the whole person, not just some functional part of the mind. They are states for which you are responsible. They are *states* rather than events or processes. All of these facts matter to the way in which you self-attribute beliefs and intentions. Because of these key commonalities between belief and intention, it makes sense to treat their epistemology together.

I have chosen to treat both belief and intention, rather than just one or another, in order to provide a generality check on the views I advance here. Too many philosophical discussions of self-knowledge focus just on belief. The danger here is that any view so narrowly focused will end up relying on special features of belief to explain how self-knowledge works, and so any such view will lack any interesting generality.

We have good reason at the outset to want some generality in our picture of self-knowledge. It is always nice to accept a simpler theory, and in this case a theory that applies to self-attribution of various states will be simpler in at least one respect than a theory that leverages different facts to explain self-knowledge of distinct states. But there are also a couple of reasons to think that self-knowledge about beliefs and intentions works in much the same way.

One important fact that unifies these states is that their self-attributions all enjoy **first-person authority**. That is: first-personal self-attributions of belief or intention are taken to be true by default, and they are also given privilege over third-personal attributions of the same attitudes, by default.

It is also plausible that self-attribution of belief and intention is **transparent** in just the same way. You can self-attribute a belief on some matter in part by making a judgment with the content of the belief to be self-attributed. When you do that, you look 'through' the question about what you believe to a question about what is true on the matter. That is why this kind of self-attribution is called "transparent." Similarly, you can self-attribute an intention on some matter in part by making a decision with the content of the intention to be self-attributed. When you do that, you look 'through' the question about what you intend to do to a question about what to do on the matter.

The fact that there are **Moorean absurdities** involving belief attribution is closely connected with the fact that first-personal belief attributions are transparent in the way I have just described. These Moorean absurdities are contents that are absurd to assert—or to judge—even though they can be true. These are of the form "*p*, but I don't believe that *p*," "*p*, but I believe it's not the case that *p*," or "it's not the case that *p*, but I believe that *p*." The possibility of Moorean absurdities for intention has been mostly overlooked in philosophical thought. It is more difficult to think of similar sentences involving intention attribution that would be absurd to assert or judge in the same way. It should not be a starting point of our study, then, that belief and intention share this further feature as well.

After further investigation (in Chapter 4) into five kinds of candidate Moorean absurdity for intention, I conclude that there are none that share all the key features of Moorean absurdities for belief. This striking fact demonstrates that transparent self-attribution can come apart from Moorean absurdities, and it suggests that intentions are not forms of belief themselves.

Nonetheless, the first-person authority and transparency that first-personal belief attributions and first-personal intention attributions share gives us reason to shape our inquiry around both kinds of attribution. Given these starting points, I hope to explain

- why we usually know what we believe and what we intend
- how we can fail to know our beliefs and intentions
- how first-personal and third-personal methods of attributing these states compare
- why self-knowledge of beliefs and intentions is important to us personally

The answers that I give to these questions set out some new forms of epistemic explanation, in applying them to the case of self-knowledge. Explanations of practical knowledge—knowledge of what you are doing, when you are doing it intentionally—can appeal to control where explanations of observational knowledge would appeal to evidence. Explanations of knowledge can also appeal to the **content plurality** of mental actions: the fact that, under distinct intentional descriptions that all apply to one mental action, that very same mental action has distinct contents. My explanation of why we value self-knowledge in a personal way also demonstrates that the value of knowledge might sometimes be emblematic rather than instrumental or intrinsic: having some knowledge of a particularly robust form can witness your possession of further virtues.

There are many more kinds of self-knowledge that are philosophically interesting than the forms of self-knowledge that I consider in this dissertation. How do you know that you exist? How do you know what you are seeing? How do you know what kind of thing you are? How do you know what would make you happy? I won't answer these questions in this discussion.

Here is a brief summary of what happens in each chapter of this dissertation.

In the Introduction, I set out the starting points for this dissertation: I aim to explain the first-person authority of belief and intention attributions by analyzing 'transparent' attributions of these attitudes made in the first person.

In Chapter 1, I set out the attitude problem for explaining these facts in the case of belief: no extant view explains how we know that we *believe* various contents, rather than taking up some other attitude to those same contents. To solve the attitude problem, I argue that you can know what kind of attitude you are taking in thought by taking it up intentionally. You can, e.g., engage in judgment intentionally.

In Chapter 2, I show how this fact contributes to an explanation of how we know what we *believe*. To provide this explanation, I introduce the related notions of embedded mental action and content plurality in mental action. The resulting view gives a full explanation of how we (usually) know what we believe, and why our knowledge has first-person authority.

In Chapter 3, I demonstrate how this view about self-knowledge of belief explains the special kind of absurdity involved in asserting or judging something of any of the

following forms: "*p*, but I don't believe that *p*," "*p*, but I believe that it's not the case that *p*," or "it's not the case that *p*, but I believe that *p*."

In Chapter 4, I generalize the view I have developed thus far to explain transparent self-knowledge of intention. I also go on the search for Moorean absurdities for intention attribution. I conclude that there are none, strictly speaking, and that this suggests that intentions are not forms of belief.

In Chapter 5, I argue that the groundlessness that first-personal attitude attributions of belief and intention enjoy on this view is not special to the first-person case. There are some cases of third-personal belief and intention attributions that must be groundless too.

In Chapter 6, I show that complete and diachronic transparent self-knowledge of your own attitudes has what I call 'emblematic value': having that kind of transparent self-knowledge is valuable insofar as it also involves having other valuable things. In particular, having complete transparent self-knowledge implies that you have no hidden attitudes, and having diachronic transparent self-knowledge implies that you have a unified self over time. Having a unified self is an important precondition on authenticity.

In the Conclusion, I summarize the main conclusions and innovations of this dissertation. I also emphasize the key role of mental action in our self-knowledge.

Chapter 1. Controlling an Attitude Problem

When you knowledgeably self-attribute a belief, you know, of some particular content, that it is something you believe. But those who aim to explain how you know what you believe often fail to explain how you know that you *believe* it. The challenge of saying exactly how you know that the attitude you take to some content is an attitude of *belief*—rather than one of intention or desire or any other attitude—is **the attitude problem.**

The solution I provide to the attitude problem is simple. I argue that you can control the kind of attitude you take up in conscious thought. When you do that intentionally, you know which attitude you are taking. This explains how you know your judgments and beliefs as such.

This chapter has three sections. Section 1 shows that several influential views of self-knowledge fail to solve the attitude problem. Section 2 analyzes explanations of knowledge in which control plays the role of justification. Section 3 solves the attitude problem in terms of intentional control over the attitude you are taking up in conscious thought.

1. The attitude problem

One reason that self-knowledge has enjoyed enduring philosophical attention is that it enjoys **first-person authority**, of the kind detailed earlier (Introduction). You are the highest authority, for example, on what you believe. Your knowledge on this topic is fantastically reliable. It's better than anyone else's knowledge of your beliefs. People come to you to settle questions about what you believe. They accept your word on the matter without asking how you know about it. A full explanation of self-knowledge of belief must explain these facts about its authority.

In this paper I focus on self-knowledge of one particular kind: knowledge of your own judgments as such. By 'judgments,' I mean the mental events with propositional contents that are (to borrow terms from Nishi Shah and David Velleman) both normatively and descriptively guided by truth.[1]

Take any **authoritative** self-attribution of belief.[2] In making that self-attribution, you know of some content (*de re*) that you believe it. For example, if you self-attribute the belief that apples are fruit, you know, of the content *apples are fruit*, that you believe it. To give this piece of knowledge a full epistemic explanation, we must explain both how you specify the content (here, *apples are fruit*) and how you know you take *that* attitude (here, belief) towards it. Those are two potentially distinct explanatory tasks.

This distinction at the level of philosophical explanation maps onto a distinction at the level of your knowledge itself. For instance, it is possible (though certainly not normal) to have a content in mind without knowing which attitude you take towards it.[3] Here is an example. Say that the thought *my son will never be good at math* pops into your head. You might be troubled by that and arrest your train of thought. You might ask

[1] Shah and Velleman (2005).

[2] Any authoritative self-attribution of an attitude is, by definition, a piece of self-*knowledge*.

[3] Compare Silins (2012) on "blurting out" that *p* in thought (p.308). I don't agree with Silins that this sort of example implies that judgment is not sufficient for belief.

yourself: "do I really *believe* that, or was I just entertaining the possibility?" I'll take this question at face value: I think it is possible to genuinely *judge* that and not know you did.

It is also possible (though certainly not normal) to have some attitude in mind without knowing which content it has. Here's an example. Say that you set your car keys down in an odd location—say, by your coffee machine. As you do that, you think to yourself, "Let me not forget: I put my car keys by the coffee machine." Ten minutes later, as you head out the front door, you can't find your car keys. You recall that you formed a belief about their location, but you can't remember what it was. Here you have an attitude in mind (a belief, specifically about your keys' location) but you don't know what its content is.

Cases of authoritative self-knowledge aren't like these. When you authoritatively self-attribute some belief, you know of some content (*de re*) that you believe it.[4] But we need to be careful to explain both aspects of this knowledge: how you know *which* content you believe, and how you know you *believe* it. It is the second part of this task that creates the **attitude problem** for many attempted explanations of self-knowledge.

Dretske saw the importance of this distinction for explanations of self-knowledge, but he saw no way to solve the attitude problem. As a result, he said that you have authority about *what* you think, but not the fact *that* you think it.[5] He reasoned as follows. You have special first-personal access to the contents of your conscious thoughts just in virtue of your thinking them, but that is not awareness of those contents *as* contents of your thoughts per se. It is awareness *with* the contents of those thoughts—e.g. awareness that *apples are fruit*. But how, then, do you know that you think these things? There is no aspect of a thought that reveals that. There is no special first-personal sign you have of your thoughts as such. You must know *that* you think in the same way you know others think: by observation. Thus, you lack first-person authority about your thoughts *as such*.[6]

Dretske drew the wrong conclusion with the right distinction. We do have first-person authority—not only about our thoughts as such, but also about our beliefs as such. But the availability of Dretske's lopsided doubt demonstrates the difficulty of solving the attitude problem, for belief as for thought. Many views have failed to solve it.

I'll discuss just a few of these views now. On one level, they are quite different. They use distinct kinds of claims in their explanations of authoritative self-knowledge. Russell talks about acquaintance, Evans about transparency of belief, and Byrne about inference. But the differences between these views belie their deeper similarity. They all try to explain how you authoritatively self-attribute *beliefs*, as such, using resources that only really illuminate how you specify the *contents* of those beliefs.

1.1. Russell

Russell claims that we are *acquainted* with "the events which happen in our minds."[7] This point doesn't apply directly to belief, as belief is a mental state, not an event. But it would help us explain self-knowledge of belief to see how you know about the events that happen in your mind—and especially how you know about your own *judgments*. Judgments are events, and they share an attitudinal aspect with belief. On Russell's

[4] I take this as definitional of 'authoritative self-attribution of belief' in this discussion.
[5] Dretske (2003a, 2003b, 2012a, 2012b). Here, "think" means *judge* (see e.g. 2012b, p.155).
[6] Gopnik (1993) argues for the same point in a different way—by using developmental data.
[7] Russell (1912), p.73. See also Russell (1910/11).

picture, you are acquainted with your judgments, if not directly with your beliefs.

However, in trying to use acquaintance to explain how you know your judgments *as such*, we would face a dilemma. We could simply stipulate that acquaintance with judgments guarantees knowledge of them *as* judgments, but we would thereby fail to explain that fact.[8] To explain it, we would have to fill out the notion of acquaintance.

Russell did further characterize acquaintance as a direct, *de re* relation of awareness that grounds all knowledge of truths. Intuitively, it is "the converse of the relation of … presentation."[9] You are acquainted with everything that is mentally presented to you.

But this characterization of acquaintance does not imply that you are acquainted with your judgments in any way that reveals their attitudinal aspect. A thought can be a judgment even if all that is presented to you, in making it, is the *content* you judge. You can judge that *apples are fruit* while thinking of nothing but apples being fruit. Moreover, it seemed in an example above that you can judge something—e.g. *my son will never be good at math*—and then reasonably wonder whether you really thought it was true in that moment. If acquaintance guaranteed awareness of the attitudinal aspect of judgments, that would not be possible. Russell's view would make that wonderment absurd.[10]

On either horn of the dilemma, a Russellian acquaintance view would fail to solve the attitude problem. Perhaps you are acquainted with the 'presented' contents of your judgments. But it does not seem that you are acquainted with their attitudinal aspects at all. Acquaintance is more naturally a relation borne to contents and objects of thought.[11]

1.2. Evans

However, we can and should still accept that judgment is essential to our self-knowledge of belief. In one short and influential passage, Evans made this point almost undeniable. He is thus credited with discovering the transparency of belief: the fact that a belief that *p* can be authoritatively self-attributed by way of judging that *p*.[12] As he put it:

> 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?' I get myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p* … If a judging subject applies this procedure, then necessarily he will gain knowledge of one of his own mental states: even the most determined sceptic cannot find here a gap in which to insert

[8] Pitt (2004), for example, simply assumes acquaintance with attitudinal aspects of thought.
[9] Russell (1910/11), p.108.
[10] Also: for Russell, acquaintance never involves *predicative* awareness. But the awareness to be explained predicates, of some thought, that it is a judgment.
[11] Contemporary ideas of acquaintance (e.g. Chalmers 2003, Horgan and Kriegel 2007, Gertler 2012) may help explain how you know the contents of your experiences nonetheless.
[12] Moran (2001) points out that Edgley (1969) made this observation first.

his knife.[13]

It is a remarkable fact that the judgment you make to self-attribute a belief can be fully "directed outward." It need not be a judgment *about* your belief, or about you, at all. The content of that judgment just is the content of the belief you self-attribute by making it.

This passage describes a method you can use to specify the *content* of a belief you have. You figure out *what* you believe on some matter—here, *p* or *not p*—"by putting into operation whatever procedure [you] have for answering the question whether *p*." But Evans's observation does not constitute a full explanation of first-person authority about belief. In fact, it makes the attitude problem look even more puzzling than before. Evans's suggestion is that you come to awareness of your beliefs as such *by way of* judgments that have contents that don't have to do with belief at all. How can that be?

1.3. Byrne

Alex Byrne has seen that the transparency of belief itself needs to be explained. In trying to explain it, he gives a proposal that is meant to solve the attitude problem in full.[14]

Byrne argues that you can actually derive an attitude from its content: you can make a genuine inference from *p* to *I believe that p*. It isn't a deductive, inductive, or abductive inference, but this inference is still "strongly self-verifying:" even when your judgment *p* is false, *I believe that p* is still true, as judgment is sufficient for belief. Moreover, even if that strong self-verification is not sufficient for knowledge, the conclusion will nonetheless have "epistemic merit," merit that enjoys only in the first person. That, Byrne argues, is enough to make sense of first-person authority about belief.

However, this limited epistemic merit will only attach to those inferential conclusions performed in a doxastic context, rather than a suppositional one. You cannot infer, just under the *supposition* that *p*, that you believe that *p*. If you could, you could conclude for any *p* that: *if p, then I believe that p*. That would absurdly imply you believe all truths.

That means that Byrne's view depends on these inferences' restriction to doxastic contexts—that is, precisely those contexts in which you really do take the attitude (belief) that you come to self-attribute in this way. Even if this sensitivity to attitudinal context doesn't require explicit awareness of the context as such, it is uncomfortable to leave this sensitivity unexplained in an explanation of self-knowledge of belief. It is particularly uncomfortable when there is an explanation in the offing—which, I will argue, there is.

1.4. A diagnosis of difficulty

This survey of attempts to explain authoritative self-knowledge of belief is far from exhaustive. But it does demonstrate the recalcitrance of the attitude problem.

The views mentioned fail to solve the attitude problem because they all focus on features connected to the *contents* of the beliefs you can authoritatively self-attribute. Russell repurposed a relation we bear to contents in order to explain awareness of attitudes. Evans emphasized the outward gaze we take in making judgments about the

[13] Evans (1982), p.225. For more on transparency of belief and self-knowledge, see Moran (2001, 2003, 2004), Barnett (2015), Williams (2014), and Shoemaker (1995, 1998, 2003).
[14] Byrne (2005, 2011, 2012). He derives the main point from Gallois (1996).

world. Byrne took self-attributions of belief to be derivable from their contents. But no feature of the content of any judgment or belief can illuminate its attitudinal aspect. The attitudes we take towards contents do not show up for us in their contents.

The problem is that it's not clear what could justify the self-attribution of a particular attitude—here, an attitude of judgment or belief. No feature of the content of any judgment or belief seems to warrant self-attribution of that particular kind of attitude. Nor must the attitudinal aspect of a judgment, or of a belief, itself be presented to its subject. It is not clear, then, how to *justify* a self-attribution of a judgment or a belief as such.

Fortunately, we don't need justification to explain self-knowledge in these cases. There is a way of explaining some knowledge—practical knowledge—that appeals not to justification, but to *control*. It is this kind of explanation that will solve the attitude problem. Before I apply it for that purpose, though, I'll introduce it more generally.

2. Control and practical knowledge

Practical knowledge, as first introduced by G.E.M. Anscombe, is the non-observational contemporaneous knowledge you have of your own intentional actions.[15] Anscombe thought there was a necessary connection between intentional action and practical knowledge: you have practical knowledge of each of your intentional actions.

I will argue that this is not a necessary condition. I'll demonstrate as much to show how there is substantive room for epistemic *explanation* in cases of genuine practical knowledge. There's an additional substantive necessary condition on practical knowledge that has not been previously recognized. This is a control condition. To show why some piece of practical knowledge really is *knowledge*, you must cite control.

Although explanations of knowledge in terms of control may sound unusual, their general form is familiar—not from philosophy, but from everyday life. To introduce the role of control in explanations of practical knowledge, let's begin with an example.

2.1. The firework show

Every year, Amy and her friend Cai watch the Fourth of July firework show from a hill overlooking their town.[16] Each time, Cai perfectly predicts which fireworks will go off, exactly when they will go off, and what colors and patterns they will display.

Amy has no idea how Cai pulls off this feat. She goes on a secret quest to find out how he does it. She learns about all the different kinds of fireworks. She watches videos on YouTube. She keeps herself updated on advances and trends in pyrotechnics. But nothing tells her how to predict the fireworks as well as Cai does, every single year.

The explanation of Cai's success is simple. His knowledge is not knowledge of independent matters of fact that he perfectly predicts. He knows the firework show because he *produces* the firework show. He designs the shows each year, oversees the setup, and uses an app to direct the release of the fireworks remotely.

---

[15] Anscombe (1957). Cf. Moran (2004b), Ford, Hornsby, and Stoutland (2014), Setiya (2008, 2009), Paul (2009a, 2009b), Schwenkler (2012, 2015), Velleman (1989).

[16] Cai is named after great pyrotechnic artist Cai Guo-Qiang. A video of one of his most remarkable exhibitions of fireworks can be found at https://www.youtube.com/watch?v=rHLd-QIb2_U. Thanks to Susanna Siegel for bringing Cai's work to my attention.

The timed releases of the fireworks are intentional actions that Cai performs. The fact that *he's doing it all* helps to demystify the knowledge he has of the fireworks show—in a way that respects the fact that it is genuine knowledge. But there's another crucial fact that matters to the explanation of Cai's knowledge: he is *in control*. If Amy learned that Cai was intentionally directing the release of the fireworks, but the fireworks were only tenuously connected to his direction, his knowledge would still be mysterious. Note that the revelation of control does not *dissolve* Amy's original genuinely epistemic puzzlement. Instead, it provides an answer to the question "How does he know?"

Because Cai exercises his control over the fireworks in remote intentional action, Amy's attempts to explain and duplicate Cai's expertise were misguided. Cai was not reading signals or interpreting patterns at all, so her efforts to find those would not help her out. There were no features of the fireworks themselves that could tell Amy when and how they would be used in the show. There's no hidden *justification* for Cai's predictions that Amy could find, even in principle.

The explanation of Cai's knowledge is an epistemic explanation that appeals to control where we might otherwise appeal to *justification*. It is an example of a general form of epistemic explanation that I will use to solve the attitude problem. For now, I'll step back to introduce the idea of *practical knowledge*.

2.2. Practical knowledge

Anscombe famously argued for a necessary connection between intentionally doing something and knowing what you are doing in doing it.[17] Anscombe individuated intentional actions as those to which a certain kind of 'Why?' question has application. The relevant kind of 'Why?' question ("Why are you doing that?") is a request for reasons that recommend the action in question, under some relevant description.

Anscombe brought out the connection between action and knowledge by arguing that this question can be rejected as inapplicable by citing a lack of knowledge. When you say "I didn't know I was doing that" in response to "Why are you doing that?" you reject the applicability of that question to your situation. The lack of knowledge seems to imply the inapplicability of the question. But intentional actions are just those to which this question is applicable. The lack of knowledge implies that the action is not intentional. That seems to imply that acting intentionally must involve knowing what you are doing.

This knowledge is not based in *observation* of your doing what you are doing. You can also reject the relevant "Why?" question by saying "I knew that I was doing that, but only by seeing myself do it," for instance. The non-observational knowledge of what you are doing intentionally Anscombe called "practical knowledge."[18]

Actions are only intentional under certain descriptions, and those are the descriptions under which you know what you are doing. Ascriptions of intentional actions are *intensional*: you cannot substitute co-referring terms in some such ascription.[19] Here is an example. Let's say that you throw away an old manuscript without knowing that it is the

[17] Anscombe (1957). See also Hampshire (1959) and Olson (1969), who make the same claim.
[18] For the purposes of this paper I adopt Anscombe's terminology for the contemporaneous knowledge you have of *some* of your intentional actions, while crucially disagreeing with Anscombe that all intentional action involves practical knowledge of what you are doing.
[19] Also see Davidson (1967/2001).

last copy of a lost Shakespearean sonnet. Your action is intentional under the description *throwing away some papers*, but not under the description *throwing away the last copy of a lost Shakespearean sonnet*—even though you are doing both. You also have practical knowledge of what you are doing under the first description, but not the second. This is the picture of knowledge in intentional action that Anscombe offered.

It is true that *when* you have practical knowledge of an intentional action, you know what you are doing under the same description under which it is *intentional*—and you do not know that based on observation. But it is not true that *all* intentional action entails practical knowledge. You can do something intentionally without *knowing* what you are doing—at least, without knowing it non-observationally. I'll make space for this alternative conception by suggesting some modifications to Anscombe's view.

2.3 Breaking the bond

Anscombe's imagined conversation centering around the 'Why?' question, does not provide real support for the necessary connection between intentional action and practical knowledge.[20] Here, I'll argue for a reinterpretation of this key thought experiment.

Note first that you can also reject the relevant 'Why?' question by saying "I didn't think I was doing that," or "I thought I was doing that, just based on my observations." This suggests that *thinking* of what you are doing in some way is required for intentional action, but practical knowledge may not be.[21]

But why, then, could you reject the 'Why?' question by citing lack of knowledge? That doesn't imply that you weren't *thinking* of your action in the relevant way (and not just based on observation). Strictly speaking, that implication does not hold. But in context, the presuppositions of the conversational context produce a condition in which it makes sense for you to express the fact that you weren't *thinking* of your action in some way by saying you didn't *know* that your action was an action of the relevant kind.

In context, the questioner asks why you are doing something that you really are doing—say, throwing away that lost sonnet. The question itself, in this case, takes as a presupposition that the action really is an action of throwing away that lost sonnet. Taking the description at face value, and deferentially accepting that presupposition, you might naturally *now* say you know you were throwing away that lost sonnet. It then expresses a natural contrast to say that you *didn't know* you were doing that in doing it.

However, not *all* ways of failing to know what you are doing non-observationally will be cases in which your own action is not an intentional one under some particular description. Consider trying to reject the 'Why?' question by protesting that you didn't

[20] This is not the only way in which Anscombe motivates the necessary connection. Another point of hers, borrowed from Aquinas, is that practical knowledge is the "cause of what it understands" (p.87). As Moran (2004b) and Schwenkler (2012, 2015) have pointed out, this is best interpreted as a claim about a formal, not efficient, cause. The claim is that an action would not be the specific action it is without the agent thinking of it in a particular way. This part of Anscombe's motivation for the necessary connection between practical knowledge and intentional action also falls short of a demonstration of that connection, for reasons parallel to the ones discussed in text.

[21] Exactly what this "thinking" comes to is a matter of dispute in itself. See Setiya (2008, 2009) and Paul (2009b) for discussion of ways to weaken of Anscombe's connection between action and practical knowledge. This detail does not substantially affect our discussion here.

know you were throwing away the sonnet, and then being asked: "In what way did you fail to know?" If you said "I didn't think of my action in that way," or "I didn't *realize* that's what I was doing," the 'Why?' really doesn't apply. But if you insisted that you thought of your action in that way (and not only based on your observations), but you just didn't meet the standards for *knowing* what you were doing, your questioner would have reason to be annoyed. That is no way to reject the 'Why?' question as inapplicable.

This reinterpretation of Anscombe's argument breaks the strongest bond between intentional action and practical knowledge. It clears the way to motivate a substantive additional necessary condition on practical knowledge: having control over what you do.

2.4. Intentional action without practical knowledge

Sometimes, you can perform an action, and perform it intentionally, without meeting all the conditions necessary to have *practical knowledge* of your very own action.

Here's an example of intentional action without practical knowledge. I have only middling control over my shots in tennis. Say I want to hit a cross-court lob into the back right court. Let's say that I focus hard, and try to hit that shot, and get a little bit lucky as I do. The cross-court lob lands in the back right. This rarely happens for me.

I propose that this is a case in which I perform an action intentionally without having knowing what I am doing *just* in doing it. I *intentionally* hit a cross-court lob into the back right. But I don't know that I'm successful except in part by relying on my eyes. I have to wait and see where the ball goes. Because my knowledge of my hitting the cross-court lob into the back right is observational, it is not practical knowledge.

Consider how different things would be for Serena Williams. She need not rely on observation at all to know what kind of shot she's hitting as she hits it. When *she* performs the same action intentionally, she knows what she's doing. The difference between us is a matter of control. That explains our difference in practical knowledge.

Here's a different kind of example. This is a case in which you lack sufficient control over the success of your attempt, even though it is not due to any failure of skill or competence on your part. Say you always get cash from a lone ATM on your corner. You are perfectly skilled in this respect. One day you come out and see that five more ATMs have been put up. You cannot tell which is the old one, and which are the new ones.[22] Unbeknownst to you, credit card skimmers have been set up in all the new ATMs, and none of them are actually connected to your bank. You can use your card in any one of them to successfully receive your ordinary amount—say, $40—but that action will not constitute a successful *withdrawal* of $40 from your own bank account.

Since you can't tell which ATM is new, you just pick one. You luckily happen to pick the real, original ATM, which really is connected to your bank. You successfully withdraw $40 in cash from your account. But it does not seem that you *know* what you are doing under that description. The threat of credit card skimming makes your success a mere accident in the face of genuine threats. You are not in full *control* of whether you successfully withdraw the cash, even though you happen to do so, in this case. In very close possible worlds, you fail to withdraw anything from your own account.

[22] This example can be seen as presenting a practical-knowledge variant on the threat present in Fake Barn Country. For discussion see, e.g., Chapter 3 of Nozick (1981), or Kripke (2011), p.166ff. The ultimate attribution is due to Alvin Goldman (1976).

What these examples bring out is the modal nature of control. Having control over something isn't just a matter of its actually going the way you intend it to go. Having control entails (but is not just) the fact that, in similar counterfactual scenarios, you could have and—given the same intentions and motivations—would have brought about the same effect. This is a distinct condition than the counterfactuals involved in the fact that you *cause* any event that you bring about intentionally. It takes more to have control over some event's happening than it does to cause that event.[23]

It is worth pausing to note that your control also limits the intentions you can form, and limits even more stringently the intentions you can *rationally* form.[24] It is absurd to buy a lottery ticket out of a billion with the express intention of winning, given that it is so utterly out of your control which ticket actually wins the lottery.[25] The control condition on practical knowledge, as a *substantive* condition on practical knowledge that is sometimes unmet in cases of genuine intentional action, is consistent with these further control conditions on intention itself. What I am proposing is that the conditions placed by control on practical knowledge are more demanding than those placed on intention.[26]

2.5. Control and justification

Now we can state the analogy I'm drawing in a more rigorous way. When you intentionally Φ, you both think of what you're doing as Φ-ing and you Φ. But in some such cases you can fail to have practical knowledge: non-observational knowledge of your Φ-ing as you Φ. It is a substantive additional necessary condition on such practical knowledge of your actual Φ-ing that you have at least a certain amount of control over your succeeding in your attempt to Φ—that is, control over your actually Φ-ing.[27]

Here is the way I'm thinking of justification that enables the instructive analogy. In some cases you believe (or judge) that *p*, and *p* is true, but you do not know that *p*. It is a substantive additional necessary condition on empirical knowledge that you have at least a certain level of justification for your belief (or judgment) that *p*.[28]

---

[23] For a discussion of the relevant counterfactuals, and a discussion of cases in which causation does not entail counterfactual dependence, see Hall (2004).

[24] Compare Shepherd (2014). This point is related to Bratman's (1999) distinct claim that intention is an attitude that controls future conduct. Intention could hardly be that kind of attitude if you could intend anything whatsoever, even that over which you have absolutely no control.

[25] Similarly, the non-voluntarism of intention (its not being under your control) seems to rule out certain combinations of intentions—see Kavka's (1983) famous case of the toxin puzzle.

[26] Anscombe (1957) recognized that something we might call 'practical knowledge' in an everyday sense—something like "knowing one's way about"—is entailed by intentional action (pp.88-89) What she didn't see was that the control required to know one's way about, and thus successfully perform an intentional action, is less control than is required to have practical knowledge in her stricter sense—in the same sense in which I am using it in this paper.

[27] Here I intend the vagueness of "a certain amount of control." The amount of control necessary will likely vary by context in ways similar to knowledge itself, and that's just as we should expect, if control is a necessary condition on practical knowledge. See, e.g., DeRose (1999).

[28] There may be additional conditions on knowledge as well, e.g. sensitivity or safety. These conditions can also be usefully analogized to control in practical knowledge. Sufficient control ensures that conditions directly adapted from sensitivity and safety apply in the case of practical knowledge. Consider sensitivity first (see e.g. Nozick 1981). If you have sufficient control over

Control is a substantive necessary condition on practical knowledge, and justification a substantive necessary condition on empirical knowledge. Each condition is dissociable from the relevant belief, and from its truth. Just as you can have true empirical belief without having empirical knowledge, you can have what we might call 'true practical belief' in intentional action without having practical knowledge. What does this imply about our knowledge explanations? An appeal to control or to justification is at least necessary—and is *sometimes* sufficient—to explain the relevant kind of knowledge.

This is the basic skeleton of the analogy. But there is much more to the analogy than just this. Here are some further ways in which control is like justification.

The first and most important further analogy is that being justified and having control both reduce the accidentality, or luck, of the belief in question. Justification does this by relating you to a preexisting state of affairs, allowing you to fit your belief to the world; control does this by bringing about the state of affairs you intend to bring about, thus allowing you to fit the world to your belief.

Another commonality is that both justification and control are gradable. You can have more or less justification, and more or less control. This makes available, for both practical and empirical knowledge, context-sensitivity of knowledge.[29] Just as you can be truly said to know an empirical fact in one context and not in another with higher standards for justification, you can be truly said to *know what you're doing* in a context with lower standards for control, and not in a context with higher standards for control.

The last two pieces of this analogy, which I present below, have a special shape. They take this form: there is a non-trivial philosophical question that characterizes the domain of justification that also applies—as a question of that kind—in the case of control as it relates to practical knowledge. The fact that these questions are theoretically open for control in a way that they are theoretically open for justification thus further supports the analogy between control and justification as they relate to knowledge of different kinds.

The first question is one about the nature of control. Is having control an *internal* fact about you or is it an *external* fact about the world? This question is one that mirrors the longstanding question about internalism and externalism about justification in empirical knowledge. On the side of internalism about control, you might point out that control involves *knowing how* to do things or a capacity to adjust flexibly to changing circumstances in a way that manifests self-aware practical intelligence.[30] In favor of externalism about control, you could argue that control ineluctably requires cooperation of conditions of the world, or that exercising control need not require any reflective access to your means—or to those very conditions that enable your continued control over some matter.[31] None of these sketched considerations are yet conclusive in this

---

whether you are Φ-ing, in Φ-ing intentionally, then the following is true: *if* you were not Φ-ing, you would not think you were Φ-ing in this way (without observation). Now consider the safety condition (see e.g. Sosa 1999). If you have sufficient control over whether you are Φ-ing, in Φ-ing intentionally, then the following is true: *if* you thought you were Φ-ing in this way, you would indeed be Φ-ing intentionally. Something like safety and sensitivity hold for practical knowledge.

[29] Compare, e.g., DeRose (1999).

[30] Cf. Setiya (2008, 2009) on know-how and intentional action. See Paul (2009b) for objections.

[31] A reliabilist about justification like Goldman (1979) might be particularly attracted to an externalist view of control on which it's reliability of successful intentional action that is required for practical knowledge in intentional action.

debate. I mean here just to make plausible that there are considerations to be adduced on either side of the debate for control, as there are in the debate about justification.

The second question is about difficult corner cases in which you (a) have control over some matter, and (b) you have true practical belief about your intentional action, but you nonetheless lack practical knowledge. Just as with Gettier problems, there will turn out to be such cases where having a true belief *and* meeting this substantive additional necessary condition on knowledge—the justification condition, for empirical knowledge, or the control condition for practical knowledge—will not be sufficient for knowledge. The inescapability of these gaps in Gettier cases derives from the dissociability of justification and truth; in the case of practical knowledge, its inescapability derives from the gap between successful intentional action and control.[32] It goes some way towards closing the gap to say that practical knowledge requires an intentional action to be the right kind of actual exercise of the control in question.[33] However, even this stipulation leaves daylight between success and control, and double accidents can occasion success in suspect ways that seem to rule out the possibility of knowledge.

There is more work to be done on the nature of control and its role in supporting practical knowledge.[34] Here I mean only to motivate a broad structural analogy between control and justification that makes available a new kind of epistemic explanation.

Now I will provide a single-case proof of concept for this proposal. The best way to see how this new kind of explanation of knowledge works is by way of example.

3. Controlling the attitude problem

Here's where we left things with the attitude problem. It seems that conscious, occurrent judgments will help us understand how we self-attribute beliefs. But there is no feature of the *content* of a judgment that reveals its attitudinal aspect, and you need not be *presented* in thought with anything more than its content to count as making a judgment. What, then, could serve as justification for your self-attribution of a judgment at all?

I have argued in the previous section that not all knowledge explanations need to appeal to justification. Some explanations—explanations of practical knowledge—appeal to control where we would otherwise appeal to justification. We need one of *these* to explain your epistemic relationship to the attitudes you take up in thought.

Let's begin by noting the sense in which a judgment can be an intentional action.

---

[32] See Gettier (1963) and Zagzebski (1994).

[33] This amendment would, for instance, rule out some cases of deviant causal chains—including Davidson's (1973/2001a) famous example of the nervous rock climber—as threats to the joint sufficiency of true belief with control for practical knowledge. Note that not all deviant causal chain examples will count as Gettier-style cases for practical knowledge, but some will.

[34] The most important question I leave aside here is whether control simply counts *as* justification of a certain kind. This proposal would have some affinities to Velleman's (1989) proposal that intentional action involves belief that the agent can reasonably expect *will* be evidentially supported once the action is accomplished. There are other questions here too. For example: is there a generality problem for control as there is for reliabilism (Conee and Feldman 1998)?

3.1. Intentional judgment

Judgment shares an attitudinal aspect with belief, but they differ in metaphysical type. To believe something is, roughly, to hold it true. To judge something is to do the same, but in occurrent thought. Beliefs are mental states, and judgments are mental events.

To be more precise: beliefs are individuated from other mental states, and judgments from other mental events, in the same way. To quote Nishi Shah and David Velleman: beliefs are just those mental states that are "governed, both normatively and descriptively, by the standard of truth." Judgments are the mental *events* that are so governed.[35]

That governance claim involves two important observations. First, judgments are in fact regulated by truth. For example, when I gain conclusive evidence that some p is false and recognize that evidence for what it is, I will (generally) not judge that p.[36] Second, judgments are normatively assessed relative to the standard of truth.

Our everyday language marks judgments in a few distinct ways. Whenever you realize that *p*, or figure out that *p* (to take two examples), you are *judging* that *p*.[37]

Because judgments share beliefs' attitudinal aspect, understanding awareness of the attitudinal aspect of judgment will help us along the way to solving the attitude problem for belief as well. You have awareness of that attitudinal aspect in *intentional judgment*.

In what sense can judgment be intentional? Judgment can be intentional insofar as you can set out to judge some things, rather than, say, imagine some things.[38] That is what you do when you set out to determine what's true. You can also set out to make a judgment about some topic *T* or a judgment *whether p*.

We need not accept voluntarism about judgment to see that judgments can be intentional.[39] Though judgment can be intentional, it is not possible to decide at will the precise content of one's judgments: you cannot, without regard for the truth of some proposition *p*, will yourself into a judgment that *p*. You don't control *what* you judge in that way, although you can control *whether* you are judging.

---

[35] Shah and Velleman (2005), p.499. They never quite apply the same terminology to the individuation of judgment, though see pp.503-5. Note, too, that they do disagree with my claim (to come) that judgment is sufficient for contemporaneous belief (p.507ff.). Compare Shah (2003), Boghossian (2003), and Gibbard (2003).

[36] This is meant to be a fairly weak restriction on the actual causal dispositions of beliefs and of judgments; see Shah and Velleman (2005), p.499. Errors must be possible, too (see Shah 2003). Cf. Davidson's (1973/2001a, 1974/2001) claim that interpretation of another as a believer involves attributing true beliefs to her. Davidson's point is holistic and Shah's is individualistic.

[37] Frege notably said judgment is the "recognition of the truth of a thought" (1979, p.294). Others say that judgment is formation of belief (e.g Crane 2001). But not all judgment is recognition of the truth of a thought: some judgments are false and recognition of truth is factive. You can also judge what you already believe, and that is not formation of belief.

[38] There may be an unavoidably Kantian flavor to any discussion of intentional judgment. But it is not obvious that Kant thought of the capacity to judge as a capacity of the *person* in the first instance. It is important to the way that I (and others) think of this capacity that it *is* exercised at the person level. Similar considerations caution against identifying practical knowledge of intentional judgments (a proper subset of judgments) with apperception (which Kant connects with all judgment)—but the similarities are tempting. Cf. Longuenesse (2001), Kitcher (2011).

[39] See, e.g., Dorsch (2009). For arguments against doxastic voluntarism, see Williams (1976) and Shah and Velleman (2005).

When you judge that *p* intentionally in this way, it is your practical knowledge that explains how you know the attitude you take towards *p*. You do not need to determine something about the content in order to know that you *judge* that *p*. You know that you are judging in virtue of judging intentionally with control over your judging as such. *p* is just what you end up judging, consciously. That is how you know that you judge that *p*. To know that is already to know your thought that *p* has a particular attitudinal aspect.[40]

Note that this account of intentional judgment does not rule out the possibility of uncertainty about attitudinal aspect in passive cases. When you are judging, but not intentionally, you may not know what attitude you are taking in conscious thought.

Note also that you need not understand the word "judge" at all in order to judge intentionally. You can hold yourself to the standard of truth, and thus engage in judgment, under a number of distinct, but equivalent, descriptions. You can think of what you are doing as *figuring out what's true*, for example. What is important is just that you think of what you are doing in a way that distinguishes it from other forms of thought that are not judgment—e.g. imagining, or supposing, or deciding what to do. Your intention must be *contrastive* in order to capture the attitudinal aspect of judgment.

3.2. Judgment and control

The upshot of the earlier discussion about control and practical knowledge was that you can perform an action—like a judgment—intentionally without knowing what you are doing. To know what you are doing in judging intentionally, you need to exercise the right kind of control. Fortunately, you have **strong control** over your judging: even just *trying* to engage in judgment ensures that you do. Thus, engaging in intentional judgment in a way that exercises such control ensures knowledge of what you are doing.[41]

To see how your attempts to judge guarantee the right attitudinal aspect, let's return to the individuation of judgment given above. You succeed in engaging in *judgment* insofar as your thought is both normatively and descriptively governed by truth.

Seeing how an attempt to judge guarantees the normative condition is straightforward. Insofar as you try to hold yourself to the standard of truth, your thoughts performed with that aim are normatively assessable by that standard.

Seeing how the descriptive condition is met is more difficult. It is more difficult to see this partly because you might think that you can just fail—by brute error, perhaps, or by inattention—to track truth in the way that seems to be demanded by the descriptive condition. You can perform some arithmetic wrong, or forget a premise you know to be true that would help you perform a better inference. In that case, would your attempt to *judge* (rather than engage in any other kind of thought) fail?

It would not, because the descriptive condition is best understood in fairly weak and holistic terms. Judgments can be incorrect and still be judgments. The normative standard would not really make sense if that were not the case. But the normative standard would equally not make sense if we were truly terrible at tracking the truth in general. What matters in any case of judgment is that you are exercising a *general* capacity that, on some suitably *general* level of description, succeeds in tracking the truth. What you do in

[40] Compare Soteriou (2013), p.316.

[41] That is the case even if you are aiming at a particular judgment, e.g. a judgment whether *p*. A failure to judge whether *p* is not a failure to judge *per se*.

aiming at the truth in general needs to be good enough.

The question of precisely what is good enough in this context is a deep one, and I cannot do justice to it here. But I will make one substantive suggestion here that serves our purposes just fine. My suggestion is that the good-enough condition in question here must be *no stronger than* the parallel condition on having the concept JUDGMENT. What that means is that you could never make a brute error in *thinking* you are judging now (when doing so intentionally) just because you are not good enough, in general, at tracking the truth when you try to do so.

Insofar as you are thinking of yourself as judging, you must have the relevant concept JUDGMENT. You wouldn't be able to have that concept unless your attempts to track the truth—to which you apply that concept, JUDGMENT—met some sufficiently stringent success condition. If you were someone who tried to judge but *in trying* to do that fairly often just happened to rehearse song lyrics or imagine outlandish scenarios, we wouldn't pity you as someone who really thinks she is judging while she is doing something else. Instead, we would pity you as someone who doesn't know what judging is at all. But if you managed to track the truth well enough, we would say you'd got it: you know what judgment is, and so you know what you are doing when you are trying to *judge* per se.

You must meet the descriptive condition, when you try to *judge*, since your trying to judge involves having a concept that guaranteed you already meet a descriptive condition of the same strength.[42] Since your attempt guarantees that you meet the normative condition as well, your attempt guarantees that you really are taking the attitude that you mean to be taking in thought. You have strong control over whether you are judging.[43]

Conclusion

Sometimes when we explain why some belief counts as knowledge we need to appeal to the justification a subject has for believing it. Other times, we need to appeal to *control* to explain knowledge—that is, in cases of practical knowledge in particular.

This chapter has demonstrated that appeal to control can solve the attitude problem: it can show how you know what kind of attitude you are taking up in thought. Other views about self-knowledge of belief, including those advanced by Bertrand Russell, Gareth Evans, and Alex Byrne, cannot solve the attitude problem because they do not appeal to control in mental action as I have done here.

It now remains to be shown how control in intentional judgment can contribute to a total epistemic explanation of self-attribution of belief. That is the focus of the following chapter, where I argue that an intentional judgment that *p* can be numerically one and the same event as an intentional judgment that *I believe that p*. In order to show how this is possible, I introduce the linked ideas of **embedded mental action** and **content plurality**.

[42] The same points apply, *mutatis mutandis*, to intentional judgments whose aims are formulated with distinct but equivalent concepts, e.g. the concept of FIGURING OUT.

[43] The same sorts of points apply for other attitudes you can take up in thought. You can control whether you are imagining, or making decisions, or deciding what to do.

Chapter 2. Embedded Mental Action and Content Plurality

You can self-attribute a belief that *p* transparently—that is, partly by way of judging that *p*. In Chapter 1, I argued that intentional judgment with control must be involved here, to make sense of the epistemology behind this transparent method of self-attributing a belief. But it is one thing to judge that *p*, and another thing to believe that *p*. What, then, is the relationship between judgment that *p* and self-attribution of a belief that *p*?

In the context of an intentional attempt to self-attribute a belief that *p*, an intentional judgment can *also be* a self-attribution of the belief that *p*. To show how that is possible, I'll introduce the linked notions of **embedded mental action** and **content plurality**.[44]

This paper has five sections. In Section 1, I introduce these two new notions. In Section 2, I apply them to help explain what happens when you transparently self-attribute a belief. In Section 3, I show that this ensures that transparent self-attributions of belief manifest authoritative knowledge of what you believe. In Section 4, I reply to a few objections. In Section 5, I compare the view I am advancing with a few other attempts to make sense of transparent self-attribution of belief in terms of agency.

1. Action and content

Mental actions are things one does mentally. Imagining what your wedding will be like can be a mental action. Recalling what your doctor told you can be a mental action. Supposing that *x=4* can be a mental action.[45] You can perform mental actions intentionally, too: you can imagine what your wedding will be like intentionally, or recall what your doctor told you intentionally, or suppose that *x=4* intentionally.

There are (at least) two different ways of individuating mental actions. First, mental actions can be individuated by their attitudinal types: *imagining* that you won gold in the hundred-meter dash is different from *recalling* that. Second, mental actions of the same attitudinal type can be individuated by their contents: judging that you won *gold* is a different action than judging that you won *silver*. Here I will take a type of mental action *simpliciter* to be a class of actions as individuated in both ways. An example of a relevant type of mental action is *judging that you've won gold*.

Here's an abstract definition of **embedded mental action**: an embedded mental action is any intentional mental action of some type *T* that also belongs to another type *U* in virtue of your having antecedently conceptualized its content in some particular way (*de dicto*) in your intention to perform that action. In the context of an ongoing purposeful mental task, when you think of the contents of some of your upcoming thoughts (perhaps just those meeting a description *d*) as being *F*, in performing each such mental action (that meets *d*) you already take that particular content to be *F*. No further move is required for you to take those contents to be *F*.

This phenomenon is not particularly circumscribed to mental actions rather than actions in general. Compare the following case in non-mental mental action: conceptualizing what you're about to do as aiming for target A partly makes it the case

[44] This section is largely drawn from my previous work in A. Peacocke (2017).

[45] For my understanding of mental action I am indebted quite generally to O'Brien (2007), O'Brien and Soteriou (2009), Soteriou (2013), Peacocke (2008), and Ryle (1971a, b, c).

that what you go on to do, when you let fly, is attempt to hit target A. Similarly, thinking of the scarf you're knitting as a gift for me is what makes it the case that making that scarf is also an action of making my gift. Here, we are particularly concerned with embedded *mental* action, but it is worth noting that embedded action is more general.

Here is an example of an embedded mental action. Imagine you are writing a story about the actors Lupita Nyong'o and Idris Elba. Let's say you need to think of something Lupita could say to Idris in response to a compliment. To do that, you can call to mind various sentences, e.g. "You flatter me!" In entertaining each one, you are not just thinking of a sentence. You are also thinking of something Lupita could say to Idris. It is your antecedent understanding of such sentences as things Lupita might say to Idris—as well as your implicit recognition that thinking of sentences is apt for the task at hand—that makes each of these acts more than just an act of thinking of a sentence. You need do no more, after entertaining each sentence, in order to think of *that* sentence as something Lupita could say to Idris. You have already built that into what you were doing in the first place. In this case, the consideration of a particular sentence is a mental action that is *embedded* into the task of thinking of something that Lupita could say to Idris. Each act of thinking of a sentence is one and the same as an act of thinking of something that Lupita could say to Idris—in particular, the sentence you are considering.

It's not only the fact that you understand what you're doing in a certain way that makes the embedded mental action take on this additional identity. It also has to be the case that performing the embedded mental action (successfully) is *actually* a way of doing the task you are setting out to perform—that is, the task in which your embedded mental action *is embedded*.[46] This fact places constraints on both the type of mental action you must choose to perform in order to carry out the broader task, and the type of contents that the relevantly embedded mental actions must have.

For example, in order to think of something Lupita could say to Idris, you could not simply imagine what various fruits look like. Imagining the looks of various fruits is not, in fact, a way of thinking of something someone could say to someone else. The kind of contents that these putatively embedded mental actions have is not the right kind of content, although the *kind* of embedded mental action you're performing in more general terms—i.e. imagining—would be appropriate to the overarching task if your imaginings had different contents, i.e. the contents that sentences have.

Similarly, if you wanted to call to mind something you know about Lincoln, you could not choose to accomplish that task simply by *supposing* various things to be true of Lincoln. You would have to engage in embedded mental actions that are *judgments* in order to call to mind anything that you know about Lincoln. Otherwise, there would be a mismatch in *kind* of mental action here (although the relevant contents, which are here propositions, would be appropriate even with embedded mental actions of suppositions).

There is another fact associated with these sorts of constraints that will be important for our purposes. There are mental tasks that others are engaged in that *you* cannot accomplish by performing your own mental actions, and vice versa. You cannot embed your own mental actions in another's mental task, for the simple reason that *your* doing something is not a way of someone else's doing something at all (and vice versa). This will become relevant in explaining the asymmetry between first-person and third-personal attributions of belief.

---

[46] Thanks to Michael Bratman for emphasizing this point.

When you do accomplish an overarching mental task by way of an embedded mental action whose performance *is* a way of completing the greater task, that individual action has the feature of **content plurality**. Under two (or more) intentional descriptions, both (or all) of which equally apply to it, that one mental action has distinct contents. Return to the case of Lupita and Idris. When you think of a sentence, your *embedded* mental action just has the content of that sentence itself, e.g. "You flatter me!" But in this intentional context, you are thinking of a sentence in order to judge of some sentence that Lupita could say it to Idris. The very same mental event is also an action of judging that Lupita could say "You flatter me!" to Idris.

On this view, to ask "what content does *that* mental action have?" is not yet to ask a question with a unique answer. In certain circumstances, it's also not enough to individuate an answer to ask "what content does *that* judgment have?"[47] This is a key implication of the content plurality of some embedded mental actions.

2. Embedded judgment in transparent self-attribution of belief

A judgment that *p*, in the context of an ongoing activity undertaken to self-attribute a belief whether *p*, can be an embedded mental action. If you are intentionally engaging in judgment (not necessarily in those terms) in order to self-attribute a belief whether *p*, a judgment that *p* can *also be* a self-attribution of a belief that *p*. You need do no more to self-attribute the belief than judge that *p*, as long as that judgment is embedded in the context of the overarching activity aiming at self-attribution of a belief whether *p*.

When you self-attribute a belief as Evans described, you need not perform any further actions after judging that *p* in order to self-attribute the belief that *p*. Here is all you need to do. First, you intentionally set out to self-attribute a belief meeting some content requirement (e.g. a belief about some topic, or a belief whether *p*). You do that by intentionally setting out to make a judgment with some content that meets the relevant requirement, already understanding the content of the judgment to come as sharing the content of a belief you have.[48] As a result of that, you actually do make a judgment with a content that meets the requirement on the belief to be self-attributed—and in so doing, you judge, first-personally, that you have a belief with that content.

The judgment that you go on to make that meets the content requirement you set is the embedded mental action here. Your antecedent understanding of the content of this judgment as the content of a belief of yours makes it the case that you already self-attribute a belief with the same content just in making this judgment. The relevant mental action has content plurality: under one intentional description (a judgment about what you believe), it has the content "I believe that *p*;" under another intentional description (a judgment about whether *p*), it has the content *p*.

Doing this requires that you implicitly understand that judgments are the right sorts of mental actions to perform in order to get at your beliefs on the matter. Because you have the concept BELIEF, you recognize that you couldn't just imagine things for the same purpose. Your recognition that you must use judgment stems from implicit conceptual understanding that the attitude you take in judging is the same sort of attitude you have as

[47] Here I assume that you can ostend a judgment with reference to some feature other than its content—e.g. by referring to the judgment you made at exactly *that* point in time.

[48] Again, you need not think of all this explicitly in terms of judgment.

when you have a belief about the world. You want to end up with a judgment about what you believe, and you can't willfully deceive yourself into thinking that the contents of some mental actions that aren't judgments (e.g. imaginings) are also the contents of your beliefs. Thus you're constrained to make judgments to self-attribute a belief whether *p*.[49]

Do we still need to provide a separate explanation of your understanding that the kind of attitude you take in intentional judgment is just that attitude you take in belief? We do not. You would not really have the concept BELIEF if (a) you had strong intentional control over whether you were judging but (b) you could not see that doing *this* sort of thing—which is to say, judgment—is the one that shares an attitudinal aspect of belief.[50] To have the concept BELIEF at all is also to understand that the sort of intentional mental action to perform in order to self-attribute belief is the mental action of judgment. We could not understand transparent self-attribution of belief as a piece of genuine self-knowledge if you did not see the appropriateness of judgment in making a self-attribution of belief. But you cannot fail to understand that if you understand what belief is at all.

3. Knowledge and authority

I have now laid out all that we need to explain how we know what we believe. Now I'll move to explain, more slowly, why self-attributions of belief made with embedded intentional judgments manifest authoritative knowledge. For brevity, and in deference to Evans's observation, I call these self-attributions "transparent self-attributions" below.

The explanation to come has three parts. I first explain why transparent self-attributions of belief must all be true. Then I explain why they must all be warranted, in un-Gettierizable ways.[51] Finally, I explain why we have first-person authority with respect to our beliefs—i.e. why these transparent self-attributions are authoritative.

3.1. Truth

Transparent self-attributions—that is, those self-attributions of belief that *p* that are made in judging that *p*—are guaranteed to be true because judgment is sufficient for contemporaneous belief.[52] Judging that *p* must involve, at the very least, having a *momentary* belief that *p*, since belief just is the mental state "that is governed, both normatively and descriptively, by the standard of truth," and judgment just is the mental action governed in precisely the same way.[53] The two share an attitudinal aspect. For that reason, one cannot perform the action without being in the corresponding state. The ontological difference between action and state does not block the implication, although

[49] Compare Shah (2003), who says there is a "*prescription* to believe that *p* only if *p* is true that frames an agent's deliberation about *whether to believe that p*," and that "accepting this prescription is one of the conditions for possessing the concept of belief" (pp.448-9).
[50] In previous work (2017) I give an extended thought example to motivate this point.
[51] Gettier (1963) showed that warranted true belief is not always knowledge. To say some warrant is not Gettierizable is just to show that the warrant in question *is* sufficient for knowledge.
[52] I do not, however, endorse the claim that judgment at some time t is sufficient for belief at any other time *t'*, or for any interval of time *T*.
[53] Quotations are from Shah and Velleman (2005). Note that I do not share Soteriou's (2013) worries about the possibility of momentary mental states.

it does block its converse.[54]

Some philosophers have argued that you can judge that *p* without contemporaneously having the belief that *p*. Arguing for this position often involves putting forward putative examples of judgment that *p* without contemporaneous belief that *p*. On further analysis of the actual examples advanced for this cause, however, we can come to see that these examples may be misdescribed. I'll discuss two such examples here, and then I will provide some more general reasons to think that no example will show that judgment that *p* is not sufficient for contemporaneous belief that *p*.

Nicolas Silins (2012) presents an example of an "accidental" judgment, a "performance error which fails to reflect an underlying belief ... you 'blur[t] out' that p, either in speech or merely in thought, consciously endorsing the proposition that p, yet failing to have a standing belief that p."[55] Silins here relies on the distinction between what he calls "standing belief" or "underlying belief"—a belief state one is in for some more extended interval—and momentary belief.[56] Silins's case seems rather to be a case of rapid doxastic change, or a case in which the agent doesn't really ever judge that *p*.

Christopher Peacocke suggests that "someone may judge that undergraduate degrees from countries other than her own are of an equal standard to her own, and excellent reasons may be operative in her assertions to that effect. All the same, it may be quite clear, in decisions she makes on hiring, or in making recommendations, that she does not really have this belief at all" (p. 90).[57] Once again the possibility in question trades on different timescales for belief and judgment. Peacocke illustrates that you might be best described as lacking a particular belief over an extended period of time even though you make genuine judgments at moments during that interval with the corresponding content. Even if that were true, it would be odd to insist that at no point during this interval, not even during those moments of judgment, does one have the relevant belief. One cannot fail, at the moment of judgment, to have the corresponding belief, although considerations about what it is to have a belief over some extended interval might bring us to admit that one doesn't "really" have the belief during that entire interval of time.[58]

Further resistance to the claim that judgment is sufficient for belief may derive from a

---

[54] Boyle (2009a) has made this point in a powerful way.

[55] Silins (2012), p.308.

[56] There is no in-principle limitation on how short-lived genuine beliefs can be. There is no absurdity in saying "I really believed that for just one moment." Consider the following example. In a hurry to catch a flight, I rush through airport security and pause, uncertain which gate is mine. I glance at my boarding pass and see "34B." I start towards gate 34B, before realizing, just one moment later, that "34B" is my seat and my gate is instead 11B. I pivot on my heel and take off in the opposite direction. In this situation, it is true that I believed that my gate was 34B—my taking a particular directed action to move towards the higher-numbered gates illustrates that—but I believed it just momentarily.

[57] C. Peacocke (1998), p.90.

[58] It's not even obvious that we must deny that one has the relevant belief over the extended interval. This scenario might best be understood as a case of conflicting belief instead. If one judges that *p* at t, one must also believe that *p* at t. But one may also judge that *p* at t while believing that it's not the case that *p* as well. For more on the possibility of having contradictory beliefs, see Chapter 3, which is about Moorean absurdities about belief.

metaphysical view of belief on which it is essentially a complex dispositional state.[59] If we accept that belief is a complex structure of dispositions towards behavior, reasoning, and mental phenomenology, then judgment may not be sufficient for belief. A judgment that *p* may sometimes manifest the dispositional structure that is belief, but at other times that judgment might occur without the presence of the corresponding dispositional state.

I will briefly argue that we should accept that judgment is sufficient for belief even if we endorse a dispositional theory of belief.

The main condition of adequacy on any view of belief—dispositional or not—is that it capture the sense in which belief is a state of taking to be true (to put it roughly). To construct a dispositional notion of belief on the general level, we might consider whether we should ordinarily include a disposition to assert that *p* in the dispositional complex that is belief that *p*. How might we decide that? We should consider whether having a disposition to assert that *p* implies (at least generically, or defeasibly) that you take it to be true that *p*, in the sense involved in a belief.

It is also the main condition of adequacy on an account of judgment that it capture the sense in which to judge that *p* is to take it to be true that *p*. We might give a dispositional account of judgment as well as a dispositional account of belief—an account on which an action's being a judgment is a matter of the dispositions it causes or manifests. If so, I cannot see any reason to label some set of dispositions that are together sufficient for judgment as insufficient for (at least momentary) belief. If both judgment and belief must capture a specific notion of taking to be true, then there should be an unbreakable implication from judgment to contemporaneous belief. This implication is entirely compatible with the possibility of judgment and lack of corresponding belief over time as well as the possibility of belief with little or no disposition towards judgment.

In principle we could also combine a dispositional analysis of belief with a different kind of individuation of judgment among mental actions. In this case, it would yet still be strange if judgment turned out to be insufficient for contemporaneous belief. The same fundamental characterization that judgment and belief share (their status as takings-to-be-true) would have to be respected in this kind of mixed analysis. On any view of the relationship between belief and judgment, then, judgment should come out to be sufficient for at least contemporaneous, momentary belief.

Those are some general reasons to think that arguments against the sufficiency of judgment for contemporaneous belief will fail. Let's return to our discussion of transparent self-attributions of belief now.

Recall that in judging that *p* in the course of the transparency method, you also at the same time self-attribute the belief for which that judgment is sufficient—namely, the belief that *p*. Thus, due to the sufficiency of judgment for contemporaneous belief, any transparent self-attribution of belief is guaranteed to be true.

3.2. Warrant

What gives any such self-attribution its epistemic right? That is, what *warrants* it?

The warrant for any transparent self-attributions of belief has three aspects: warrant for the *self*-attribution of belief that *p*; warrant for the self-attribution of *belief* that *p*; and

[59] See e.g. Schwitzgebel (2002, 2012), and Cassam (2014), pp.117-19. I'd like to thank an anonymous reviewer at *Philosophical Studies* for highlighting this point.

warrant for the self-attribution of belief *with that specific content*, *p*.[60] I'll address each aspect of the warrant for a transparent self-attribution of belief in turn.

3.2.1. *Self*-attribution

Here you are entitled to attribute the belief to *yourself*, rather than someone else, insofar as your transparent self-attribution involves no method of self-identification that could err. Transparent self-attribution of belief involves no positive identification of yourself at all. Transparent self-attributions of belief are thus immune to error through misidentification.[61]

It is not the case here that there is a particular ground that serves as justification for attributions of beliefs to yourself as opposed to others. Instead, the fact of the matter is that you cannot err in this way, and so you are *entitled* to the self-attribution in question given that you have the first-personal concept. Since judging "I believe that *p*" at all requires having that concept, anyone who actually transparently self-attributes a belief is guaranteed to have entitlement to such a *self*-attribution of belief.[62]

3.2.2. Self-attribution of *belief*

Why are you warranted in self-attributing a *belief*, as opposed to some other kind of attitude? You are warranted in the self-attribution of a *belief* for two reasons.

First, your awareness of the attitudinal aspect of your judgment is ensured by your engaging in judgment *intentionally*, and with strong control. Your understanding of belief ensures that you understand that taking up that attitude, rather than another, is what you do in belief. If you understand belief, then, you are entitled to understand what you are doing in judging intentionally as appropriate for self-attribution of belief. Putting it this way gets things round the wrong way, though. It is better to say: given your competence with the concept BELIEF, you are entitled to *judge* intentionally in order to self-attribute belief. And when you do that, your knowledge of what you are doing intentionally as self-attributing a belief enriches your target judgment that *p* into a self-aware attribution of belief that *p*.[63]

The second aspect of warrant for transparent self-attribution of *belief* in particular has to do with what is involved in having the concept BELIEF. It is trivially true that you have to have that concept to use the transparency method, since it involves thinking of things as beliefs. It is not trivial, however, that part of the understanding you must have if you have the concept BELIEF entitles you to apply that concept in the course of using the transparency method.

Part of what it is to have the concept BELIEF is to constrain yourself to use the

---

[60] I'm grateful to O'Brien (2005) for pulling apart these aspects of warrant.

[61] Shoemaker (1968), Pryor (1999). See also Wittgenstein (1958).

[62] For more on using the first-person in self-attribution of belief, see Boyle (2009b, pp.153-4).

[63] It is also worth recalling, at this point, the previous explanation of how you know that you are judging when you engage intentionally in judgment. Your strong control over whether you are judging ensures that your belief that you are judging amounts to knowledge. Here, an explanation of control steps in for an explanation of warrant. Note also that I am not here implying that you yourself need to be explicitly aware that judgment is sufficient for belief (in those terms).

mental action that is judgment (perhaps not thus conceptualized, by you) in using the transparency method. That implies that you would not use other mental actions in the same context, and that—given certain idealized conditions, other concepts, and capacity for more sophisticated reflection—you would reject as inappropriate the use of any other mental action in the place of judgment. It is this aspect of having the concept BELIEF that makes it the case that you are entitled to apply the concept in the transparency method. Since the transparency method requires the concept BELIEF, any user of that method is thus entitled to self-attribute a belief.[64]

To see why it is important to have this additional conceptual entitlement in warranting a self-attribution of a belief, we can consider what it would look like if someone lacking the concept BELIEF—and thus lacking the implicit understanding that entitles its application in the transparency method—tried to use the transparency method. Consider an agent called "Erraticus." Suppose that Erraticus has the same control over his mental actions, and the same practical knowledge that that control implies, as the rest of us do. Now further suppose that Erraticus often attempts to use the transparency method for belief and fails, because he does not engage in judgment as the relevant embedded mental action. For example: sometimes, when asked for his belief whether p, he'll start making suppositions, come to a supposition that p, and say: "I believe that p."

Erraticus could, in fact, stumble on the right way to use the transparency method as a matter of mere accident. In one instance, he might actually try to self-attribute what he calls a belief by way of making judgments. But even if Erraticus did this, he would not be warranted in what he took to be his self-attribution of 'belief,' because he clearly does not have the concept BELIEF. His lack of conceptual understanding is revealed by his erratic attempts and failures to use the transparency method for belief.

To recognize that it is Erraticus's lacking the concept BELIEF that matters to his lack of warrant in using a pseudo-transparency method is also to see that your having the concept BELIEF matters to your having warrant in using the actual transparency method for belief. Thus an account of the warrant you have for making a self-attribution of a belief, rather than any other attitude, must make essential reference to the entitlement that your conceptual understanding bestows on your use of the transparency method.

Together practical knowledge (due in part to control) and conceptual entitlement account for the warrant involved in transparent self-attribution of *belief* in particular.

3.2.3. Self-attribution of a belief that *p*

Finally, you are warranted for self-attributing a belief *with the specific content p* because to judge that *p* intentionally involves the content *p*'s being consciously thought. To judge intentionally is also to judge consciously, such that judging that *p* (when *p* is true) is a way of being occurrently aware that *p*.

A transparent self-attribution of a belief that *p* rather than, say, a belief that *q* is warranted due to the consciousness of the contents of intentional judgments. When your judgment that *p* is an intentional mental action, your consciousness has (at least in part)

[64] Here I do not mean to endorse the general principle that any inference or application of a concept whose availability to the subject is required for possessing that concept are inferences or applications to which the possessor is then entitled at any time. In its general form, this principle has interesting counterexamples. See Boghossian and Williamson (2003) for extended discussion.

the content *p*. Such consciousness with the content *p* that you have in making that judgment is all that is needed by way of warranting a transparent self-attribution of a belief that p rather than any other belief.[65]

This warrant should not be seen as evidential or inferential justification. The point is not that some relation of yours to your own consciousness that *p* (in judging that *p*) provides support for your self-attribution of the belief that *p*.

One more general point is important here, and it applies more generally to the warrant involved in transparent self-attribution of belief. When you transparently self-attribute a belief, your warrant cannot be 'Gettierized.'[66] There is no question of your having the wrong sort of warrant, or deriving your self-attribution from false lemmas, and so on. Competent intentional judgment that *p*, performed with the intention of self-attributing a belief, builds in all the warrant you need for that self-attribution that *p*—which is the same mental action as the judgment that *p*.

3.3. First-person authority

Why does first-person authority attach to such transparent self-attributions of belief? Recall that ordinary first-person authority involves a default presumption of truth, and a default privilege over third-personal attributions.

Transparent self-attributions of belief are presumed to be true because they must be: judgment is sufficient for contemporaneous belief, so a belief attributed in making a judgment with the same content must be (at least temporarily) a belief you really do have. They are accepted over others' attributions of belief to you because nobody else has such a secure way of attributing beliefs to you. Anyone else can indeed perform an intentional mental action to figure out what's true and to attribute a belief to you, but anyone else's resulting judgment that *p* will not be sufficient for *your* believing that *p*.

Moreover, nobody else can initiate your intentional mental actions, or share your knowledge of what you are doing in performing some such intentional mental action. You know what you are doing when you are doing it intentionally (and with control) precisely because it is *your* action. Recall the points made above about constraints on embedded mental action: your own mental actions cannot be embedded in others' tasks in a way that would license the use of some third-personal transparent attribution of belief.

These facts about the epistemic merits of transparent self-attribution of belief would only vindicate the presumptions involved in first-person authority attaching to *all* ordinary first-personal attributions of belief given a further important fact: we presume of each other that we make transparent self-attributions of belief by default. If we could not assume of one another that we were using this method of self-attributing beliefs, we

[65] Those who take content externalism to threaten self-knowledge (e.g. Boghossian 1989) may disagree that this is all that's necessary by way of warrant here. I take the line endorsed by Burge (1996), Heil (1988), and Peacocke (1996) on this point: there is no such threat. Those still concerned about content externalism should at least note one nice feature of any given embedded mental actions: one and the same intentional mental action cannot enjoy two distinct environments that might contribute to the individuation of content. Cf. Burge (1996) on self-verifying judgments.

[66] Cf. Gettier (1963). The un-Gettierizability of the warrant for transparent self-attributions does not depend on any one particular interpretation of what goes wrong in Gettier cases.

could not use the epistemic merits of this method to rationalize the deference we give to first-personal self-attributions of belief in general.[67]

Here is a reason to think that the explanation I've offered is a strong explanation of first-person authority. This explanation closely the authority we have in self-attributing beliefs to facts about what it *is* for some judgment or belief to be mine. The fact that nobody else can perform *my* actions—and thus nobody else can make the same kind of transparent attribution of belief to me with the same epistemic credentials—is a basic fact about what it is for some action to be mine. The fact that I have practical knowledge of those intentional actions I perform (with sufficient control) is also inseparable from their being *my* actions. A similar point applies to another aspect of the warrant you have for your transparent self-attributions of belief: nobody else is conscious of the mental events of your mind in the same way that you are. If somebody else were conscious of those events in the same way, they would not be the events of *your* mind.

The fact that this explanation of first-person authority exploits important facts about what it is for some thought or action to be yours, as opposed to anyone else's, is good reason to accept it. It demonstrates what is special about the first-personal perspective by using facts that are absolutely inseparable from the first-personal perspective.

4. Objections and replies

4.1. Content crowding

The first objection I'll consider concerns the substantive proposal made earlier in this chapter about embedded mental action and content plurality. Those mental actions that are embedded in ongoing tasks in the right way can have distinct contents under distinct intentional descriptions that apply to them. A critic might complain that this 'crowds' the content of one mental action in a metaphysically unacceptable way.[68]

The complaint cannot be that the content of these mental actions is crowded in a completely incomprehensible way, as I have offered in the exposition above a way to understand how mental actions gain content plurality, and I have described specific circumstances in which this happens. There should be nothing more mysterious about one and the same event being more than one action at once than there is anything mysterious in the fact that one's action of knitting a scarf can also be the action of making a gift.

I suspect that the metaphysical suspicion leveled at multiple identities of mental actions in particular derives from implicit dependence of an understanding of mental action as linguistic in nature. Individual sentences cannot have multiple contents, and so if thinking a thought is like slotting a sentence into a spot, a thought could not have multiple contents either—whether or not it was an action. But this is not a mandatory picture of thought, and in fact it seems to be a misguided picture of thought. A fuller discussion of this dominant metaphor of thought should wait until another time.

I should also clarify here that nothing in my view implies that *thoughts* in the abstract sense Frege meant can have multiple contents.[69] A thought in this sense is something like

[67] The same point will arise again in Chapter 6 to make sense of some ways we talk about the value of self-knowledge.
[68] Thanks to Brie Gertler for making this objection to this view (personal correspondence).
[69] Frege (1956).

a proposition. Any particular proposition has its content, and no other content, necessarily. That is what it is to be the particular proposition that it is. This fact about propositions mirrors the fact about sentences above. You might call this feature of propositions and sentences their "content singularity." It is an essential feature of them.

The content singularity of propositions is also consistent with my view. It is, in particular, mental *actions* that have content plurality. The mental context is what allows for multiplication of contents in one and the same event. We have our thoughts in intentional (and intensional) contexts, and we perform mental actions for particular intentional purposes, and that is what allows thoughts in *that* sense to have content plurality under circumstances in which one mental action is embedded in another task.

4.2. The rarity of active self-reflection

A critic might also question how often we really embark on this project of intentionally directed thought in order to probe our own beliefs. If the answer is ''not often,'' it might be thought that this account cannot do all that much to explain the full scope of privileged access, or of first-person authority. You are authoritative, the critic might claim, over far more of your own beliefs than those that you transparently self-attribute.[70]

It seems to me that we do, quite often, use the transparency method—at least as often as we explicitly consider the question of what we ourselves believe. Though its philosophical explanation is somewhat complicated, actually using this method can be so simple as to be practically mindless. It doesn't require deliberation of any sort—a judgment that p can express, rather than form, a belief—and you need not formulate anything complex to yourself in order to do it.

Yet the transparency method need not actually be used all that often in order to explain the full scope of privileged access or first-person authority. To explain your privileged access to some state such as belief, we need only to explain why you have a method you could use, at any point, to self-attribute a belief such that the self- attribution in question is more likely than any third-personal one to amount to knowledge. A similar point applies for first-person authority: to say you are an authority on what you believe is not necessarily to say that you often explicitly consider your beliefs as such. All that the claim of first-person authority implies is that, were you to consider what you believe, your word would trump anyone else's word on the matter. For these reasons, the understanding of the transparency method presented in this paper still offers a powerful way to make sense of both privileged access and first-person authority.

4.3. Diachronic belief and fallibility

The transparency method produces infallible self-attributions for the synchronic case, but judgment at some moment *t* is not sufficient for belief over any extended interval of time. That means that even transparent self-attribution of belief is not infallible when you aim to self-attribute a *diachronic* belief—one that lasts over some extended interval of time. You might worry, then, that the view I have advanced in this chapter (and the previous one) may be unable to explain authoritative self-knowledge of diachronic beliefs. You might also worry that this is most of the authoritative self-knowledge that we care about;

[70] I'm grateful to an anonymous reviewer at *Philosophical Studies* for presenting this objection.

it is a rare case when we ask ourselves (or others) what we (or they) believe right at that very moment, without caring about whether we have believed the relevant proposition for a while or whether we will continue to believe it.[71]

It is undeniable that the transparency method as explained above cannot ensure perfect knowledge of diachronic beliefs. It is important to note the potential for error in the transparency method as applied to these diachronic cases. Still, neither of these concessions implies that the transparency method can do nothing to explain self-knowledge of diachronic belief.

Importantly, the epistemology of ordinary transparent self-attribution of belief need not change substantively in cases where you *do* know what you believe over time. There may be cases in which defeaters should keep you from endorsing a momentary self-attribution of belief as capturing your diachronic state of mind. But the warrant described in this chapter seems more than sufficient to warrant diachronic self-attribution of belief as well, at least in a significant number of cases.

The authority of these self-attributions can also be retained across time. It is still the case that practical knowledge of what you are doing, and consciousness of what you are thinking, gives your transparent self-attributions of *diachronic* belief significant reliability. The third-personal methods of belief attribution anyone else might use to capture your doxastic set are still usually more fallible—in more ways, and in more circumstances—than your own transparent self-attributions of belief.

In Chapter 6 below, I return to the question of what kinds of errors (and omissions) in self-attribution of belief are available on the view I endorse here. I discuss the particularities there, as a way of understanding how your transparent self-knowledge is one among various kinds of valuable self-knowledge.

5. Other agency-based views of self-knowledge

In this final section, I respond to three other philosophical discussions that use aspects of agency to and explain authoritative self-knowledge of belief. I contrast my view with those of Richard Moran, Lucy O'Brien, and Matthew Soteriou.

5.1. Moran

Moran argues that your *epistemic* authority on the matter of what you believe is explained by your *deliberative* authority over what you believe.[72] You make up your mind about what to believe. You deliberate about what is true in order to form and reaffirm your beliefs in conscious thought. And in ordinary cases, such rational deliberation does actually determine what you believe. That must be the case, given that you are held responsible for the beliefs that you do have. Your responsibility speaks to your agency.

Moran's work is groundbreaking in the field of self-knowledge because it connects that knowledge with agency. But he has not successfully pinpointed *what* it is about being a doxastic agent that properly explains your epistemic authority about your beliefs.

Take, for instance, the claim that you make up your mind about what to believe. That

[71] I'd like to thank Peter Epstein and an audience at Harvard for expressing this point forcefully.
[72] Moran (2001). For responses see O'Brien (2003), Shoemaker (2003), Reginster (2004), Lear (2004), Heal (2004), and Wilson (2004). For Moran's replies see Moran (2003, 2004).

is true, but it does not immediately imply awareness about what you believe. You can make yourself dinner, too, but you still need to look in the oven to see whether it's done.

Take, on the other hand, the point that you are responsible for what you believe. That, too, can be true, without implying that you have epistemic authority on the matter. Parents are held broadly responsible for their children's behavior even before they are alerted to all the terrible things their children have done at daycare.[73]

In later work on this topic Moran recognized that his original discussion of deliberative agency was insufficient to answer specific epistemic questions.[74] For that reason he added another stipulation into his view. When you deliberate about what *to* believe, he said, you are entitled to assume that such deliberation really does form (or reaffirm) your beliefs. But this added stipulation faces the same sort of attitude problem that faced Byrne's inferential view. Applying this entitlement successfully already requires a certain level of sensitivity to the deliberative context. You can deliberate with suppositions, or you can deliberate with judgments. It's only in the latter, doxastic context in which you are entitled to think that your deliberations determine your beliefs.[75]

Moran comes closest to specifying the features of doxastic agency that matter to self-knowledge when he draws an analogy between belief and intentional action.[76] He draws from Anscombe and from Kant in observing that "in belief as in intentional action, the stance of the rational agent is the stance where reasons that justify are at issue."[77] Your stance towards your own beliefs is, primarily and properly, a stance in which you can justify your taking certain beliefs rather than others. Similarly, your stance towards your own actions is, primarily and properly, a stance in which you can justify taking that action rather than another. Only degenerate cases of belief or intentional action diverge.

As I argued above, Anscombe's considerations on intentional action are essential to understanding our self-knowledge of belief. And the analogy that Moran draws here does track important features that belief and intentional action share. But the most illuminating application of Anscombe's observations is not an analogy at all, but a literal application. There is intentional action in thought, and these intentional actions involve the sort of practical knowledge that Anscombe discussed. Importantly, you can engage in judgment intentionally. When you do that, you know what you are doing as such. That is how you know the type of attitude you take to the contents you go on to consider in this process.

Moran missed the fact that intentional action plays a straightforward role in explaining how you know your own beliefs. He related *belief* and intentional action in an analogy instead of seeing that engaging in judgment can just be an intentional action.

5.2. O'Brien

O'Brien recognizes that the primary difficulty facing accounts of authoritative self-knowledge is a problem in accounting for knowledge of what she calls the 'force,' and I

[73] Similar points about responsibility apply to other 'agentialist' positions (as labeled by Gertler (forthcoming)): see Boyle (2009b, 2011a) and Bilgrami (1998). Compare Reginster (2004).
[74] Moran (2003, 2004).
[75] Moran recognizes this limitation on the entitlement itself, but does not explain it.
[76] Moran (2001), pp.124-8.
[77] Moran (2001), p.127.

have called the "attitudinal aspect," of some authoritatively self-attributed thought.[78] She recognizes that the case of judgment is particularly central. But O'Brien's original explanation in terms of agency actually presupposes, rather than explains, the knowledge in question. She takes it that a subject has awareness of what she is doing in thought when she "realis[es] the practically known possibility of judging that *P*."[79] For that to *be* a practically *known* possibility at all, we need to have some understanding of why it is in the subject's control to judge whether *p*. However, O'Brien does not explain that.

O'Brien's view sometimes looks more puzzling than that. In an extended treatment of agency in self-knowledge, she argues that "a subject being agent aware of her [mental] action is *constituted by* the action being the product of the subject's consideration of possibilities, grasped as possibilities."[80] Applied to judgment, this claim would mean that your knowledge of what you are doing when you intentionally engage in judgment *is constituted by* that engagement in judgment being the product of your consideration of different possibilities for you—e.g. imagining, deciding what to do, etc. But that seems to be a category error. The knowledge in question cannot do the work we need it to do if it is not *constituted by* genuine awareness. But the condition provided here does not even rule out deviant causal chains of the sort that worried Davidson in intentional action—let alone ensure any awareness of the sort needed.[81]

It is also worth noting that O'Brien and I disagree about how practical knowledge in intentional judgment explains self-knowledge of belief. She is pessimistic about the possibility of explaining self-knowledge of belief in terms of self-knowledge of judgment partly because she accepts a dispositional account of belief, instead of the account I favor here, as proposed by Shah and Velleman.[82] In past work, I have argued that even dispositionalist accounts of belief should accept that judgment is sufficient for belief.[83]

5.3. Soteriou

Soteriou and I agree on a key ingredient in the correct explanation of authoritative self-knowledge of belief: the awareness each of us has in engaging in mental action intentionally.[84] Soteriou believes you can intentionally engage in judgment, and that you know what you are doing in doing that. He and I also agree that the special awareness you have of your own intentional action cannot specify more for you than the intention on which you are acting.[85] For example: if you are intentionally engaging in judgment whether *p*, your knowledge that you judge that *p* cannot be fully explained with reference to the awareness you have of what you are doing intentionally. The consciousness of the

[78] O'Brien (2003, 2005, 2007). For a response to O'Brien (2007) see Howell (2008).
[79] O'Brien (2005).
[80] O'Brien (2007), p.120, emphasis added.
[81] Davidson (1963/2001) gives an example of a rock climber who intends to let his partner fall to his death. In thinking of this possibility, the rock climber gets so nervous that his hands slip on the rope, and he inadvertently lets the partner fall to his death. This, Davidson points out, is not an intentional action, despite its being caused directly by the intention in question.
[82] See O'Brien (2005), and see Shah and Velleman (2005).
[83] See A. Peacocke (2017).
[84] Soteriou (2013).
[85] Soteriou (2013), pp.316-20.

judgment in question comes in at this point, on his view and on mine.

Unlike O'Brien, Soteriou is careful not to reduce the practical knowledge in question to something else. However, there are two key differences between my view and his.[86]

The first is that Soteriou does not recognize the importance of control to the discussion of *knowing* what you are doing in thought, and does not see that we have such strong control over the attitudes we take in thought. But without mentioning these features, our solution to the attitude problem would be incomplete. Leaving it at that would be like finishing an explanation of knowledge without mentioning justification.[87]

The second distinction is a distinction between Soteriou's explanation and my explanation of self-knowledge of belief. We disagree about the relation between a judgment that *p* (performed in ongoing intentional engagement in judgment) and the self-attribution of a belief that *p*. I have argued that the two can be one and the same. Soteriou, in contrast, argues that after intentionally engaging in judgment whether *p*,

> One believes that what one set out to do was to work out whether *p*. If one believes that one has done what one set out to do, then one will believe that in concluding that *p* one has worked out that *p*. If one believes that one has worked out whether *p*, then one believes that one knows whether *p*. On the assumption that knowledge is a state, this means that one's belief about what just happened entails a belief about one's current state, and not simply a belief about some past event. And on the assumption that the obtaining of the state of knowing that *p* entails the obtaining of the state of belief that *p*, this makes plausible the claim that in acquiring this belief about what has just happened one has acquired the belief that one believes that *p*.[88]

Soteriou grounds your self-attribution of a belief that *p* in your belief that you *know p*, which itself is meant to be derived from your just having consciously worked out that *p*.

It is itself controversial whether consciously working out that *p* implies that you believe that you *know* that *p*. Take, for example, a case in which you have just learned how to perform some new arithmetic function—e.g. multiplication. You might work out that *three times twelve is thirty-six* without thereby believing you *know* that.

Even if we set aside this matter, Soteriou's view has unfortunate consequences for the authority with which we self-attribute beliefs. On this view, your knowledge that you believe that *p* can only be as well warranted as your knowledge that you *know* that *p*. That does not square with the fact that everyone has first-person authority about belief—

---

[86] Actually, we also disagree about some related metaphysics. Soteriou (2013) identifies a need for a category of an 'occurrent mental state' (e.g. a belief that one is judging that *p*) that is constitutively dependent on some temporally extended mental event (here, the judgment that *p*) (see especially pp.246-8). I think the invention of the category confuses the issue. He might have put the same points in terms of being *in* a state for a temporally extended period. His use of the same ad hoc notion also requires judgment to be a temporally extended event, but any particular judgment that *p* occurs all at a moment. Restricting oneself *to* judgment can be extended.

[87] In a discussion of knowledge in cases of intention-in-action, Soteriou (2013) writes: "you are apprised of much of what is going on in your mental life because it was your idea to begin with" (p.321). That explanation needs to be supplemented with an explanation of control.

[88] Soteriou (2013), p.352.

even those who admit that they are not sure whether they *know* those propositions that they *do* at least know they believe. Soteriou's view thus also fails to vindicate the common intuition that authoritative self-knowledge of belief is epistemically immediate.

Conclusion

In this chapter I explained how the control we have in engaging in judgment contributes to a full epistemic explanation of transparent self-attribution of belief. I introduced the connected notions of embedded mental action and content plurality to show how a judgment that *p*, when performed intentionally as a *way* of figuring out whether you believe *p*, can be one and the same action as a self-attribution of a belief that *p*. When a self-attribution of belief is made in this way—that is, transparently—it has great epistemic credentials. This way of self-attributing belief is infallible in synchronic cases (though fallible when it comes to self-attributing diachronic beliefs). The warrant any such self-attribution of belief has is especially strong, and cannot be Gettierized. The components of that warrant—which include immunity to error through misidentification, practical knowledge, and consciousness of intentional mental action—ground a full explanation of the first-person authority that self-attributions of belief enjoy.

Chapter 3. Moorean Absurdities about Belief

In the last two chapters I explained how you know, with first-person authority, what you believe. I did that by explaining how you can self-attribute beliefs transparently, and explaining why the self-attributions made in this way have the status of knowledge.

In this chapter I will step back and analyze which features of belief as a propositional attitude mattered to the epistemic explanation I gave above. I will do this for two reasons.

First, I want to demonstrate that belief's having those key features, and thus being transparently self-attributable, provides a solution to Moore's paradox.

Moore's paradox arises from the fact that there are contents that are absurd to assert or to judge, which are nonetheless satisfiable contents. These contents traditionally take the form "*p*, but I don't believe that *p*," or "*p*, but I believe that it's not the case that *p*." These contents express ways the world could be: you can fail to believe truths, and you can believe falsehoods. But it is clearly absurd to judge or assert anything with either form. The contents themselves are called "Moorean absurdities." Moore's *paradox* arises from the tension between the absurdity and the obvious satisfiability of the contents. To solve the paradox is just to explain why it is absurd to judge or to assert a content with one of these forms even though the content is satisfiable. In this paper, I do just that, by using the resources of the epistemic explanation given above, and by highlighting certain important features of the propositional attitude that is belief.

The second reason I abstract to these key features of belief is to generalize. Other mental states, or mental actions, can have the same features that both allow for transparent self-attribution and generate Moorean absurdities. At the end of this paper, I'll lay out these features in generalized form, and show how the mental act of judgment has all the necessary features. That means that judgments are transparently self-attributable, and it also means that there are Moorean absurdities involving judgment attributions (in lieu of *belief* attributions as such). But the absurdity of judging or asserting something with one of these contents is explicable in just the same way as it is explicable in the case of belief, due to the relevant shared features of judgment and belief.

The chapter following this does the same for decision and intention. There I argue that the mental acts that are decisions are transparently self-attributable, and so are the mental states that are intentions. They both have associated Moorean absurdities, and the explanation of why these contents are absurd to judge or to assert proceeds along the same lines as before.

Here is a summary. In Section 1, I'll analyze the case of belief to show which facts about belief mattered to the epistemic explanation of transparent self-attribution that I gave in previous chapters. In Section 2, I'll show how these facts explain the absurdity of judging or asserting something of the 'omissive' form "*p*, but I don't believe that *p*," or the 'commissive' form "*p*, but I believe that it's not the case that *p*."[89] In Section 3, I'll compare my solution to Moore's paradox to other proposed solutions to Moore's paradox, to demonstrate the advantages of the explanation I endorse. Then, in Section 4, I will generalize the key facts about belief that mattered to all of this discussion. There I'll apply this generalized framework to show that the mental act of judgment is transparently

[89] I'll also explain the absurdity of judging something of a 'negative commissive' form, "it's not the case that *p*, but I believe that *p*."

self-attributable and that there are Moorean absurdities for judgment that can be explained with reference to the same kinds of facts as those that apply to belief.

1. Key facts about belief

In the last chapter, I explained why you can transparently self-attribute beliefs, and why the self-attributions that you form in this way have the status of authoritative knowledge. There are six key facts about belief that mattered to these explanations. Here are these key facts about belief, with short reminders of how they mattered to these explanations:

1. There is a mental action—namely, judgment that *p*—the performance of which is sufficient for the contemporaneous existence of a belief with the same content, *p*. (This fact explains why transparent self-attributions of belief are necessarily *true* in the first person, but not in the third person.)
2. You have control over your performance of this *kind* of mental action—that is, whether you are judging. That implies that you can intentionally engage in judgment. It also implies that when you do engage in judgment intentionally, you know what you are doing *as* judging (or in equivalent terms that match those that characterize the intention on which you are acting). That involves knowing the kind of attitude you are taking up in thought. (This fact partly explains why transparent self-attributions of *belief* as such are *warranted* in the first person.)
3. The practical knowledge you have of what you are doing is already first-personal and immune to error through misidentification; it involves no positive identification of yourself among objects. (This fact explains why transparent *self*-attributions of belief are entitled—and thus warranted—in the first person.)
4. To have the concept BELIEF, you must recognize that doing *that* kind of thing in thought—judging—is the appropriate kind of mental action to use to self-attribute what you believe. (This fact also partly explains why transparent self-attributions of *belief* are warranted, and it lets you use judgment as a means to self-attribute beliefs. It makes available the embedded mental action of judging that *p*, which is, in this context, the same event as judging "I believe that *p*.")
5. The successful intentional performance of judgment that *p* involves consciousness with the content *p*. (This fact explains why transparent self-attribution of a belief with a *particular content* is warranted in the first person. It also figures into the asymmetry between the first person and the third person, as nobody else has consciousness of your judgments in the same way.)
6. Your mental action of judging that *p* can, in the relevant intentional context, be one and the same event as that of your judging "I believe that *p*." That is, the judgment that *p* can be properly embedded into an ongoing mental task that is accomplished just by making this embedded judgment. (This fact explains why no transition is needed to warrant the ultimate transparent self-attribution of a belief that *p*. It also explains why a method of belief attribution with the same epistemic credentials is not available in the third person: nobody else can embed *your* mental actions into their mental actions in the same way, or vice versa.)

Here I mean to state necessary and sufficient conditions on belief's being transparently

self-attributable in a way that endows the resulting self-attributions with first-person authority. In the previous chapters I showed why these conditions were sufficient. A brief discussion will now show why each of these conditions is necessary. Not meeting any one of these conditions could 'break' either the knowledge status, or the first-person authority, of the kind of self-attribution of belief you might try to make by using the transparency method described in the previous chapter. Here is how the lack of any one would affect the truth, warrant, or authority of transparent self-attributions:

1. Without this sufficiency, transparent self-attributions of belief could be false at the time at which they are made.
2. Without control over the kind of mental action you are engaged in (here, judgment), you could engage in judgment intentionally but without practical knowledge of what you are doing. Since practical knowledge is an important part of the first-person authority you have in self-attributing beliefs, your authority would be affected by a lack of practical knowledge too.
3. Without first-personal immunity to error through misidentification, you might not be entitled to attribute the relevant belief to *yourself* among others—and your self-attribution would not be authoritative in this way from the first-person position.
4. Without this conceptual connection, you would not be simply entitled to use judgment to self-attribute belief; you would have to do more to connect the mental action of judgment to the self-attribution of a belief as such.
5. Without consciousness with the content of your judgment, you would have to have some other way of knowing the content of what you are thinking, and your self-attribution of a *particular* belief might not be authoritative from the first-person position.
6. Without the identity of the relevant mental actions (here, a judgment that *p* and a judgment "I believe that *p*") you would need a warranted way to move from one of these contents to another, and so more would be needed to justify your self-attribution of belief than just competence performance of self-attribution of belief that *p by way of* the judgment that *p* itself.

I return to these six conditions later in this chapter. Each one is also necessary to the explanation of Moorean absurdity for belief that I offer below. This makes sense of the intuitive connection between transparent self-attribution of belief and the absurdity of these Moorean judgments and assertions. It is because transparent self-knowledge of belief is so readily available that judgments or assertions with Moorean contents are absurd in the way that they are.

2. Moorean absurdities for belief

In this section I introduce the three forms of contents that I call "Moorean absurdities for belief," and I explain why it is absurd to make any judgment or assertion with a content of any of the relevant forms.

Traditional Moorean absurdities for belief come in just two forms, omissive (O) and commissive (C). They are contents with the following general form:

(O) *p*, but I don't believe that *p*.
(C) *p*, but I believe that it's not the case that *p*.[90]

Substitute some proposition in for *p* in either content schema, and you'll get a content that is absurd to assert or to judge, even though it can be true. The same goes for another form of Moorean absurdity I will consider in this chapter, the negative-commissive form:

(N) It's not the case that *p*, but I believe that *p*.[91]

The absurdity of such judgments and assertions is special to the first-personal, present-tense case. There is nothing at all strange, let alone absurd, in asserting "*p*, but she doesn't believe that *p*," "*p*, but she believes that it's not the case that *p*," "*p*, but I didn't believe that *p*," or "*p*, but I believed that it's not the case that *p*," and so on.[92] There is also nothing absurd about merely *supposing* (either out loud, or to yourself) something of the form of (O) or (C). You could easily entertain the proposition that something is true without your believing it, or that something you believe is false. The attitudinal aspect of the act with this content—i.e. whether it is a *judgment* rather than another mental act, or an *assertion* rather than another speech act—matters to whether or not it is absurd.

Another key feature that matters to the absurdity of these assertions and judgments is that they are conjunctive. Judging or asserting *p* and "I don't believe that *p*" separately is still strange, but understandable; the same goes for judging *p* and "I believe that it's not the case that *p*." The more mental or temporal distance there is between such judgments and assertions, the less odd the pair of acts seems. There is something absurd in judging or asserting the two conjuncts of a content with the form (O), (C), or (N) *together*.

As mentioned above, the contents that are themselves *instances* of any of these schemas (O) – (N) are known as "Moorean absurdities." What I aim to explain in this section is why it is absurd to judge or assert any of those contents. The fact that judgments and assertions with such contents are absurd is something I'll take for granted

[90] Sorensen (1988) seems to have been the first to label these content schemas "omissive" and "commissive." Note that there are other first-personal present-tense forms for Moorean absurdities, like "I can't believe it, but it's a fact!" The absurdity of judgments and assertions with some these slightly divergent forms will be explicable in the same way I explain the absurdity of assertions and judgments with contents of the form (O) or (C) above.

[91] You might take instances of (N) to be instances of (C) as well insofar as judging or asserting that *p* is tantamount to judging to asserting that it's not the case that it's not the case that *p*. However, given the opaque context of belief assertions, and issues about double negation in vagueness and indeterminacy, I will treat (C) and (N) as distinct forms here. This will come to matter more in the next chapter, when we are considering forms of Moorean absurdity involving intention attribution.

[92] Perhaps there is something odd about asserting "the dodo is extinct, but in the future I will not believe it," but certainly nothing so strange as to count as absurd. Here I leave aside the tricky issue of assertions in the first-person plural. "The dodo is extinct, but we don't believe it" may sound less strange than the corresponding first-person singular claim. That in itself is an interesting fact, if it is a fact, but I don't want to address it here. There are more basic issues that come out in consideration of the first-person singular version.

in this chapter.[93] What is at issue is *why* these judgments and assertions are absurd.

Here is how my explanation will proceed. I'll start by addressing the absurdity of making a judgment of any of the relevant forms (O) – (N) when you are engaging in judgment *intentionally*. I'll argue that it is *impossible* to make a judgment with any omissive, commissive, or negative-commissive content when you are intentionally engaging in judgment as such, and thus have practical knowledge of your judging. The reasons that it is impossible to do that are closely related to the reasons that transparent self-attributions of belief have the status of authoritative knowledge.

How about judgments that are made when you are not intentionally engaged in judgment as such, and so lack practical knowledge of what you are doing? Judgments with contents with the forms (O) – (N) are not *impossible* in these non-intentional contexts, but they are absurd by virtue of their proximity to the central intentional cases. When you are engaged in judgment, but not intentionally, it is still *irrational* to judge something with any omissive, commissive, or negative-commissive content.

What about the absurdity involved in cases of assertion, rather than judgment? Sincere assertion of *q* manifests occurrent judgment that *q*. Thus an explanation of the absurdity involved in *judging* some Moorean content will support further explanation of the absurdity involved in *asserting* some Moorean content as well.[94] There are some limited, unusual cases in which the absurdity of asserting some content of any form (O) – (N) is overridden by an understanding that the agent misspoke in some way, but these cases will not concern us much here.

That gives an overview of the structure of the explanation I'll present below. I'll also make three important points about the methodology of this discussion before I launch into it. Two points are basic facts, and one is about tests of explanatory adequacy.

The first fact on which we will rely repeatedly is that someone who makes a judgment or assertion with some content must have all the concepts involved in that content.[95] In particular, to make any judgment or assertion with a content of any of the forms (O) – (N) above you must have the concept BELIEF. Having that concept involves having various capacities to self-attribute beliefs, including the capacity to transparently self-attribute beliefs. That will be crucial to the explanation I provide below.

The second important fact is that making a judgment with a conjunctive content is to consider the matters of *both* conjuncts together, in one context. To judge anything with a content of any of the forms (O) – (N) above, then, is to have in mind the matter of whether *p* and the matter of whether you believe that *p*, all at once—not sequentially, in distinct contexts. That is not to say that you must always consider these matters together. It is just to say that when you can be truly said to make a judgment with a *conjunctive* content, you are considering the matters together. If you were considering the matters

---

[93] I *won't* take for granted that *believing* one of these things is absurd, though there are good arguments to the effect that that is absurd too; see Shoemaker (1995, 2009).

[94] I here accept Gottlob Frege's (1956) view of assertion as a "manifestation of [a] judgment" Contrast what Green and Williams (2007) call "Shoemaker's principle" (p.12). This view does not contradict the popular view that assertion is to be individuated in terms of its constitutive norms—see, e.g., MacFarlane (2011).

[95] This derives from what it is to have a concept. See, e.g., Evans (1982) on the Generality Constraint. Contrast what Green and Williams (2007) call "Searle's principle" (pp.21-22), which says that believing that *p* implies the ability to think that *p* occurrently.

separately or sequentially, it would not be right to attribute to you a conjunctive judgment. It would be more realistic to attribute to you sequential or separate judgments.

When I have completed the explanation of why it is absurd to judge or to assert something with a content of any of the forms (O) – (N), I will test the explanation against some conditions of explanatory adequacy. These conditions of adequacy derive from the basic profile of Moorean absurdities. I will test to make sure that the explanation I offer implies that the judgments and assertions in question are absurd only in the first-person, only in the present tense, and only in conjunctive form. I will also show how the explanation demonstrates the relevance of the mental context to the absurdity, since the contents so judged or asserted are not absurd on their own. Finally, I'll consider whether the explanation implies that related contents are absurd to judge or to assert. If it does, we can test our intuitions about absurdity against those implications as well.

2.1. Explaining the absurdity

Let's start, then, by considering why it is absurd to judge some content of the form (O), (C), or (N) when you are intentionally engaged in judgment as such. I'm going to argue that it is impossible to do that. I'll start by assuming that an arbitrary agent *does* do this, in order to perform a reductio on this assumption. For specificity: assume that our agent Tasha makes a judgment with a content that is an *omissive* Moorean absurdity, and that she does that in the course of engaging in judgment intentionally. For some *p*, she judges something with the form of the content "*p*, but I don't believe that *p*."

What is involved in her doing that? Since she has strong control over whether she is engaging in judgment, she has practical knowledge of the kind of mental action she is performing, i.e. judgment.[96] She must also have the concept BELIEF; otherwise she couldn't make any judgments with contents having to do with belief explicitly. Her making this conjunctive implies that she is currently considering the matter of *p*, and the matter of whether she believes that *p*, all at once.

Given that much, we should start to see that all that is needed for transparent self-attribution of belief is already in place in her making this judgment. To have the concept BELIEF is to recognize that judgment is the appropriate mental action to use to self-attribute beliefs, and since she is currently considering whether she believes that *p*, she must see that what she is currently doing intentionally—judging—is what she needs to self-attribute a belief. Her judgment that *p*, when she makes it, is also sufficient for her believing that *p* in that moment and in that context. Just in judging the first conjunct, she does all she needs to do to make a transparent self-attribution of a belief that *p*. Given that she is considering whether she believes that *p*, and intentionally engaged in judgment, and that her having the concept BELIEF links her current activity to what is required to make a self-attribution of a belief, her judgment that *p* also qualifies as a self-attribution of belief that *p*. Just in making the judgment that constitutes the first conjunct of the omissive Moorean absurdity, she *also* judges that she believes that *p*. That is, she makes an additional judgment that directly contradicts the second conjunct, that she doesn't believe that *p*.

While it is in general possible to make judgments with contents that are in fact logical

[96] See Chapter 1 for the definition of strong control and an argument that each of us has strong control in judging intentionally.

contradictions, that can only be done when the agent fails to see the contradictory nature of the content of her judgment. But here the contradictoriness of her judgment—enriched to include the double judgment involved in the first conjunct—could not be more apparent. She is said to be judging *p*, *I believe that p*, and *I don't believe that p* all at once. Since that involves such a direct contradiction, it is not possible for her to judge at all. Tasha simply cannot make the judgment that she is described as making here.

There is a technicality here that is important to note. In the previous chapter, I noted that the content plurality of the embedded judgment was dependent on the agent's engaging in judgment intentionally *in order to* self-attribute a belief. But that explicit intention is not loaded into the example we are currently considering. Here, the agent is intentionally engaged in judgment and she is considering what she believes on the matter of *p*, but she may not be intentionally engaged in judgment *whether p in order to self-attribute a belief whether p*. Does that mean that she might not, in making the judgment *p* that is the first conjunct of the omissive Moorean content, also make the judgment that she believes that *p* at the same time? If that were possible, then there might be a non-absurd way to make a judgment with the content "*p*, but I don't believe that *p*," even in the context of ongoing intentional judgment.

I think, however, that that is not possible. I'll show as much by way of an analogy.

Consider that someone is intentionally engaged in chopping wood, and that she knows that chopping wood is a way to upset her dog, who is sensitive to the noise and confused by the procedure. She might not be chopping wood *in order* to upset her dog; she may instead be chopping wood in order to make logs for her fireplace. Might she, in these circumstances, know that she is chopping wood *without* knowing that she is upsetting her dog? It seems that she can—but only if the matter of whether she is upsetting her dog is not at the forefront of her mind. She could defend herself against the accusation that she upset her dog *intentionally* by saying "I wasn't thinking of it." But if the matter of upsetting her dog is one she is currently considering, and she knows she is chopping wood, and she knows that doing that is a way to upset her dog, it seems she cannot fail to know, in chopping wood, that she is upsetting her dog.

This is analogous to the situation that might be seen to pose a problem to the explanation I have given so far. If someone is engaged in judgment intentionally, and she knows that judging that *p* is a way of self-attributing a belief that *p*, and the matter of whether she believes that *p* is something she is currently considering, she cannot fail to know, in judging that *p*, that *p* is something she believes. Only if she were not currently considering the matter of what she believes with respect to *p* could she miss the fact that her judging that *p* manifests her belief that *p*. But the very content of the omissive Moorean absurdity implies that the agent in question *is* currently considering what she believes on the matter, and her understanding of belief ensures that she recognizes that judging that *p* is a way of self-attributing a belief that *p*. That means that any context that involves intentionally engaging in judgment will be a context in which an agent *cannot* make a judgment with the omissive Moorean content "*p*, but I don't believe that *p*."

So far I have explained the absurdity of judging something with a content with the omissive form when you are intentionally engaging in judgment. But what about making a commissive or negative-commissive judgment in a context like that? The explanation in these cases cannot proceed in exactly the same way. Above, I relied on the fact that a judgment that *p*—the first conjunct of an omissive or commissive Moorean absurdity—is,

in context, also a judgment with the content "I believe that *p*," and that directly contradicts the second conjunct of an omissive Moorean absurdity. But that doesn't directly contradict the second conjunct of a commissive Moorean absurdity, which has the content "I believe that it's not the case that *p*." Similarly, even if the judgment of the first conjunct ("it's not the case that *p*") of a negative-commissive Moorean absurdity is *also* a belief attribution ("I believe that it's not the case that *p*"), that *further* judgment's content does not directly contradict the second conjunct ("I believe that *p*"). Some further fact is needed to explain what's wrong with judging something of the commissive or negative-commisive form, even in a context where you are engaging in judgment intentionally, and thus with knowledge of what you're doing.

The further fact that is needed here might seem to be the fact that you cannot judge of yourself, all at once (i.e. in conjunction), that you believe *p* and you believe that it's not the case that *p*. But you can judge that, sensibly, and truly. It is actually possible to have directly contradictory beliefs, and it is possible for you to have contradictory beliefs at one time, and it is possible for you to recognize this fact about yourself. However, reflection on what makes all of this possible will show why it is impossible to conjoin these self-attributions *specifically* in the context of commissive and negative-commissive Moorean judgments.

What makes it possible for an agent to have beliefs that directly contradict each other in their content—e.g. the belief that *p* and the belief that it's not the case that *p*? What makes that possible is that one and the same agent can inhabit distinct **rational contexts** at distinct times.[97] Rational contexts are individuated by the believer's body of evidence on some question, her patterns of attention, the accessibility of her memories, her cognitive load, her local risk aversion, her cognitive capacities and skills, and so forth. In the simplest case, when an agent gains evidence, or changes her mind about which risks are worth taking, it is easy to make sense of a change in her beliefs as well. We usually appeal to some such shift in values, memory, evidence, attention, skills, or more in order to explain why someone goes from believing one thing to believing the opposite. There is nothing mysterious or objectionable about this kind of understanding of change in belief.[98]

What we are interested in here, though, are the conditions in which an agent can be truly said to have contradictory beliefs *at the same time*. On the picture of rational contexts I am drawing, it is not possible for one agent to inhabit more than one rational context at one time. But it *is* possible for an agent to believe that *p* and believe that it's not the case that *p* at one time. That is possible because the set of rational contexts in which the agent manifests the various dispositions that flow from the belief that *p*—including the disposition to judge that *p*, to act as though *p*, and more—can fail to intersect with the set of rational contexts in which she manifests the various dispositions that flow from the other belief that it's not the case that *p*, including the disposition to judge that it's not the case that *p*, to act as though it's not the case that *p*, and more.[99]

---

[97] I mean here to establish the notion of a rational context as a context that captures all that which is *rationally* relevant to the formation of a judgment or a belief, or, more generally, to whether or not it is reasonable to take something as true.

[98] Lewis (1986), p.30ff. Also see Stalnaker (1986) Chapter 5 and Lewis (1982).

[99] In the previous chapter I rejected a dispositionalist analysis of belief. Note that nothing in this discussion requires a dispositionalist analysis of belief. What it does require is the fact that

Since an agent encompasses various dispositions to be in these distinct rational contexts, she can be truly said to have the two contradictory beliefs all at once. To say of her that she believes that *p* and that she believes that it's not the case that *p* just requires that we understand her beliefs as involving dispositions that manifest in distinct rational contexts.

Why is that the *only* way to make sense of that claim? Because it couldn't quite generally be the case that attributing a belief that *p* to a thinker leaves wide open the question of whether or not she believes its negation. If that were so, then to attribute a belief to someone would not be to convey all the information about her dispositions to act and judge and so forth that such an attribution *does* convey. Belief wouldn't be what it is—an attitude of taking something to be true, that is both descriptively and normatively governed by truth—if its ascriptions allowed this question to remain open in an absolute way.[100] So there must be some non-trivial delimitation of cases in which the attribution of contradictory beliefs makes sense.

I am here suggesting that the delimitation in question has to do with rational contexts. In particular, if we have settled what someone will and will not do, think, say, and so forth in *one* maximally specified rational context, there will be *only* one answer to the question of whether she believes *p* in that context.[101]

None of this directly implies that one cannot *self*-attribute directly contradictory beliefs, although I have been speaking of these attributions of contradictory beliefs in their third-personal versions. All that this does imply is that, in order to self-attribute contradictory beliefs, one must be implicitly thinking of oneself in different rational contexts when one manifests and thus expresses each of those beliefs. Especially when considering one's diachronic dispositions, one's tendencies towards one thing and the opposite over the course of a period of time, one may well recognize that one believes that *p* and that one believes that it's not the case that *p*. There is nothing to stop any thinker from taking the stance on herself as a person as her acquaintances might from the outside, and coming to the same conclusion: that she has contradictory beliefs. A famous example arises in David Hume's *Treatise* when he talks about his own different beliefs on philosophical matters in different rational contexts (living his ordinary life, that is, versus sitting down and thinking about abstract philosophical questions).[102]

It is, then, not always absurd to judge, or to assert, "I believe that *p* and I believe that it's not the case that *p*." It is absurd, however, in those situations in which one is thinking of oneself in one and the same rational context for the purposes of *both* belief ascriptions. It is not only absurd in those circumstances: it is impossible. To have the concept BELIEF is to understand that contradictory beliefs can be attributed to one agent only insofar as those beliefs are understood to manifest in distinct rational contexts.

Given this understanding of the situations in which you can and cannot self-ascribe directly contradictory beliefs, let's return to our example of a judgment with a content with the commissive or negative-commissive form, made in the course of engaging in judgment intentionally. For specificity, let's take an arbitrary agent, Leon, who judges for

---

believing that *p* involves or implies various dispositions. This weaker claim is implied by the view of belief that I favor, which I borrow from Shah and Velleman (2005): belief is the propositional mental attitude that is normatively and descriptively governed by truth.

[100] For more on what belief is, see Nishi Shah and David Velleman (2005).

[101] The answer could, of course, be that she has no belief; it just couldn't be that she has *both*.

[102] See, for example, 1.4.7.9ff. in the *Treatise*.

some specific *p* "*p*, but I believe that it's not the case that *p*." He judges this in the course of intentionally engaging in judgment in particular.

In this context—as argued above in Tasha's case—judging the first conjunct of this content also involves a self-attribution of a belief that *p*. That means that Leon is both self-attributing the belief that *p* and the belief in its negation. Leon could only do that as long as he thinks of these two beliefs as manifesting in distinct rational contexts.

However, there is something in the conjunctive judgment that fixes a particular rational context as the relevant one to *both* belief ascriptions: the simple judgment, that *p*. Judging that *p* here, rather than just self-attributing the belief that *p*, indicates that one is in some such particular rational context with respect to the matter of *p*. Since Leon actually makes the judgment that *p*, the natural way to understand his self-attribution in the second half of the judgment is as a self-attribution of a belief to him in that specific context. But the self-attribution made in the first conjunct of the whole judgment is also a self-attribution to him in that very context. The conjunction of these self-attributions is not possible for an agent like Leon who has the concept BELIEF.

The same explanation applies, mutatis mutandis, for the impossibility of making a negative-comissive Moorean judgment when you are engaging in judgment intentionally.

Thus far I have explained why it is not only absurd but impossible to judge something with a Moorean-absurd content—of the form of (O), (C), or (N) above—when you are intentionally engaging in judgment. What remains is to explain why it is still absurd to make any such judgment outside of this intentional context, and why it is absurd to assert something with the relevant content as well.

Once we have the core case involving intentional judgment on the table, these other less central cases can be explained in reference to it. It is the similarity between (i) a non-intentional judgment of any Moorean form and the impossible (ii) intentional judgment of any Moorean form that explains the absurdity of the former. Anyone who makes the non-intentional version of the judgment—that is, a judgment made *not* in the course of intentionally engaging in judgment as such—has all the resources she needs to avoid the impossible judgment in the intentional case. She has the concept BELIEF, and she has the capacity to exercise strong control in judgment and thus know what she is doing. She has all that she needs to self-attribute beliefs transparently, in a way that grants her authoritative knowledge of what she believes. If she makes a Moorean judgment despite the immediate availability of these resources that would grant her such authoritative knowledge, then she does something absurd. She fails to exercise her rational capacities to their best advantage, and so she is *irrational* in making that Moorean judgment. This explanation of the absurdity of non-intentional Moorean judgments applies equally well to the omissive, commissive, and negative-commissive cases.

Assertions of Moorean absurdities are likewise absurd because sincere assertion manifests occurrent judgment. Any assertion of a Moorean absurdity is itself absurd insofar as it manifests an absurd judgment.[103] This explanation of the absurdity in Moorean assertion also applies equally well to the omissive, commissive, and negative-commissive cases.

That completes the explanation of the absurdity in judging or asserting anything with a content of the form (O), (C), or (N).

[103] In cases of insincerity, the assertion in question will be unbelievable to its hearers. Compare Williams (1996).

2.2. Testing the explanation

Let's test the explanation using the tests I mentioned in the beginning of this section.

First: any good solution to Moore's paradox must explain why the mental context matters to the strangeness in judging or asserting a Moorean absurdity, as the contents themselves are not absurd at all. This has been accomplished on this explanation. It is the fact that an agent can *do* something in thought—that is, intentionally engage in judgment—that makes it absurd to judge or assert something of the form (O), (C), or (N). An agent's actual or potential practical knowledge of her mental actions is crucial here.

Second: any solution must also explain why only first-personal judgments have this absurdity. That, too, has been explained here. The kind of transparent attribution of belief that makes Moorean judgments and assertions absurd is thoroughly first-personal, as was explained in Chapter 2. Practical knowledge is knowledge of what *you yourself* are doing. Consciousness with the content of your own judgments is something only *you* have merely in virtue of judging something intentionally. And your judgment that *p* is only sufficient for *your* believing that *p* in the moment in which you judge it.

Similar facts explain why only *present-tense* Moorean judgments and assertions are absurd to make. Since a judgment that *p* is only sufficient for your believing that *p* in the same moment that you make that judgment, there is room to judge that *p* while acknowledging that you did not previously, or will not in the future, believe the same.

Finally, it is important to recognize how the conjunctive aspect of a Moorean judgment or assertion contributes to its absurdity. It would not be impossible, or absurd, for you to make separate judgments (or assertions) with the form of the individual conjuncts of (O), (C), or (N). That is because the world-directed judgment could be made outside of the context of considering what you believe on the same matter. It was crucial to the explanation above that what you believe on the matter of *p* is on your mind *as you judge p* (or, in the negative-commissive case, as you judge *it's not the case that* p).

We can also test the strength of this explanation of the absurdity of judgments of this form by looking at other implications this explanation has.

What I have said above implies, for example, that judgments or mental actions involving both a judgment that *p* and a recognition of ignorance with respect to whether or not one *believes* that *p* should also be absurd. That is, *intentionally* judging "*p*, but I have no idea whether I believe that *p*," is absurd. That is why it is more natural to capture a line of thought with this contour as having the content "it is the case that *p*—but wait, no, do I really believe that?" In this second characterization of the content of a line of thought, the change in belief is captured mid-thought by a break and an explicit refusal to identify with the original commitment that it is the case that *p*.

The explanation of absurdity involved in Moorean judgments and assertions with commissive or negative-commissive forms implies is that there is actually something *more* absurd in certain cases of judging "*p*, and I believe it's not the case that *p*" than there is in just judging "I believe that *p* and I believe that it's not the case that *p*," *even though* the first such judgment also involves a self-attribution of a belief that *p* in those cases. For precisely what makes the former absurd and the latter intelligible is that an explicit judgment that *p* makes salient the particular rational context one currently inhabits, and thus fixes the context of the attribution in the second conjunct in a way as to rule out the intelligibility of contradictory self-attributions.

This portion of the explanation also implies that certain *other* intentional judgments should not be absurd at all—in particular, any such intentional judgments whose second conjuncts are tweaked to 'escape' the implication that the belief that it's not the case that *p* is attributed to the person in the maximally specified rational context in which the first conjunct is itself judged. For example, it should seem fine to judge, intentionally, "*p*, but I sometimes believe that it's not the case that *p*," or "*p*, but I have a persistent belief that it's not the case that *p*." And these do indeed seem fine to judge, or to assert.

2.3. Taking stock

Here I have argued that it is not only absurd but *impossible* to make Moorean judgments when you are intentionally engaged in judgment. I used that fact to explain why it is absurd, even though it is possible, to make Moorean judgments or assertions when you are not engaged in judgment intentionally. It is absurd in those cases because it is irrational not to use all the resources you have readily available to self-attribute beliefs transparently, and thus authoritatively and knowledgeably.

A nice feature of the explanation I have given is that it uses barely more resources to solve Moore's paradoxes than I have used to explain why transparent self-attribution of belief generates authoritative knowledge. It thus honors the close connection between transparent self-knowledge of belief and the absurdity of Moorean judgment or assertion.

The explanation I have given uses the six key facts about belief that I glossed in Section 1 of this paper—the six facts that were both necessary and sufficient for the availability of transparent self-knowledge of belief. To leverage these six key facts in a solution to Moore's paradox, we also had to acknowledge a few further points, none of which I take to be terribly controversial. The facts I have already used are these:

7. In conjunctive judgment, you consider the matters of all conjuncts at once.
8. Directly contradictory beliefs are possible for an agent to have only if those beliefs manifest in distinct rational contexts.
9. To have the concept BELIEF, you must understand that it makes no sense to attribute directly contradictory beliefs to one person in the same rational context.

These three further facts contributed to the solution to Moore's paradox I have given.

But there is yet another fact that I have not yet appealed to that we need in order to complete the explanation of Moorean absurdity presented here. I have discussed why it is impossible to *make* a Moorean judgment in the course of engaging in judgment intentionally. That is, it is impossible to make that judgment in that context from the first-person perspective. It is similarly absurd from the first-person perspective to make the judgment non-intentionally, or to make the assertion.

But Moorean judgments and assertions do not only look absurd from the first-person perspective. It is absurd from the third-person perspective to think of someone *as* making that kind of judgment, and it would be absurd to hear someone *else* making that kind of assertion as well. What is involved in seeing that as absurd?

Importantly, it involves seeing the agent in question as having all the resources of transparent self-attribution available to her. It also involves taking the transparent method of self-attributing beliefs as the default method of self-attributing beliefs. We assume of

one another—that is, from the third-personal perspective—that we use the transparency method to self-attribute beliefs. This fact about the assumptions that we make in the third person was also important in Chapter 2 to complete the explanation of first-person authority in self-attributing beliefs. It is no less important in this solution to Moore's paradox. The same point will arise again in Chapter 6, where I address the value of transparent self-knowledge. We should add this to our list of key facts about belief:

10. We assume of one another that we make transparent self-attributions of belief by default, and only use other methods when this method is unavailable.

3. Other proposed solutions to Moore's paradox

Now that this solution to Moore's paradox has been laid out in full, I will survey a few competing solutions to Moore's paradox. I do this in order to demonstrate the distinctive advantages of the solution that I have given. Other solutions to Moore's paradox, I will argue, do not succeed in showing why it is definitively *absurd* to judge or to assert something with the form of (O), (C), or (N).

To argue for this conclusion, I will use some of the same tests I used above to assess the solution I have given to Moore's paradox. I will also distinguish between certain categories of norm violation and genuine absurdity. Not all irrationality, lack of warrant, and inconsistency qualifies to explain absurdity. We will also see that some of the proposed solutions to Moore's paradox run into the same problem that motivated Chapter 1 of this dissertation: the attitude problem, which is the problem of explaining *how* you know the attitudinal aspect of the kind of thought you are engaged in (e.g. judgment).

Let's begin by looking at the solutions to Moore's paradox given by the first philosophers to consider the question.

3.1. Moore and Wittgenstein

G.E. Moore first discussed odd sentences of the form "it's raining but I don't believe it" (an omissive form) or "I believe that he has gone out, but he has not" (a commissive form) in 1944. He explained the oddity in asserting the omissive sentences by claiming that your asserting "it's raining" does in some sense imply that you believe that it is raining, and that this would contradict what you assert in saying you don't believe it.[104] He explains the oddity in asserting the commissive sentences by claiming that your asserting "he has not gone out" does in some sense imply that you don't believe that it's *not* the case that he has *not* gone out—i.e. that you don't believe that he has gone out. The contradiction between what you imply in actually asserting the world-directed conjunct and what you actually assert in the other conjunct, the self-attributive one, is what generates the absurdity involved in asserting a sentence of either form.

Wittgenstein, who was taken with this discovery made by Moore, offered a very similar explanation of the oddity in asserting either of these things, but with a kind of implication running in the other direction.[105] The claim "I believe that *p*" is used much like the claim that *p*, at least in their assertions. You can, in fact, say that you believe *p* in

[104] See Moore (1993) and discussion in Green and Williams (2007), Introduction.
[105] See, e.g., Wittgenstein (1980), and discussion in Green and Williams (2007), Introduction.

order to communicate to someone that *p* is in fact the case. That means that you approach a contradiction when you assert "*p*, but I believe it's not the case that *p*," because the second conjunct will communicate to your hearer something that directly contradicts the content of your assertion of the first conjunct.[106]

These foundational attempts to explain the absurdity in making Moorean assertions are suggestive, but neither attempt is successful in providing a full solution to Moore's paradox. Neither attempt explains the absurdity in *judging* something of the form of (O), (C), or (N) listed above. It is not entirely clear that Moore or Wittgenstein even saw the absurdity in the case of judgment. They focused their discussions mostly on assertive utterances, and used Moorean absurdities to reflect on the nature of assertion itself.

This will not do to explain the absurdity in Moorean judgment. If the absurdity arose just from an implication in communication, there might be nothing at all absurd in quiet judgment of a Moorean absurdity. The agent doing the judging might simply miss the implication; in fact, the implication might not even be present in a non-communicative context like that of silent judgment.

For the same reasons, other proposed solutions to Moore's paradox that rely for their explanations solely on features of assertion, or the communicative context of utterance, will not suffice as full explanations of Moorean absurdity in judgment.[107] These include views that appeal to facts about pragmatics in assertion and those that mention the self-defeating nature of Moorean assertions.[108]

As Green and Williams have rightly noted, there is an asymmetry between those views that aim to solve Moore's paradox just with facts about assertion and those views that aim to solve it just with facts about judgment. Since assertion manifests judgment, facts about judgment can explain absurdity in assertion. But not all facts about assertion apply to judgment in the same way.

3.2. Contradictory beliefs

A more promising line of thought appeals to facts about the implications of Moorean judgments in order to solve Moore's paradox. Various philosophers have suggested that an agent's making a Moorean judgment implies that she has contradictory beliefs.[109]

This fact alone does not suffice to explain the full absurdity of judging or asserting something of the form of (O), (C), or (N). As discussed above, an agent can have contradictory beliefs, as long as they manifest only in distinct rational contexts. It is certainly not an enviable position. To have contradictory beliefs must be a violation of epistemic norms of some kind or another (no matter what kind precisely is at issue). But mere norm violation, in a way that is psychologically possible for ordinary agents, is not enough to demonstrate *absurdity*.

To provide a full explanation of the absurdity in judging or asserting a Moorean content, then, an account of this form will need to add something to this claim in order to

[106] It is a little more difficult to extract from Wittgenstein's remarks a viable explanation of the oddity in the omissive case, but this is not the most serious problem with Wittgenstein's explanation.
[107] See, e.g., Searle and Vanderveken (1985).
[108] See Green and Williams (2007), Introduction.
[109] See Green and Williams (2007) pp.10-11, Heal (1994), Sorensen (1988), and Hintikka (1962).

show why this case of having contradictory beliefs is especially absurd. In order to fill the gap, Green and Williams, for example, suggest that a "minimum of reflection" suffices to demonstrate to the agent that her judgment of some Moorean content implies that she must have contradictory beliefs.

There is a real problem with this amendment to the account. The first is that it encounters the attitude problem all over again. Green and Williams suggest that the implication of contradictory beliefs arises from the fact of the agent's *judging* some Moorean content, not merely from the content itself. But in order to use "a minimum of reflection" to recognize this fact, that minimum of reflection must yield to the agent a recognition that she has *judged* the relevant content. Without a solution to the attitude problem which demonstrates that it really is minimal reflection that is needed to recognize the attitudinal aspect of your thought, we should be suspicious of the claim that minimal reflection is required to recognize the absurdity of your own judgment here.[110]

What this suggests is that a full solution to Moore's paradox cannot be presented with absolutely no reference to a theory of self-attribution of beliefs. Solving the attitude problem is a matter of getting a better understanding of how we understand our own thoughts and the attitudes that they manifest. If a solution to the attitude problem is needed for a solution to Moore's paradox, then self-knowledge must be relevant to the solution of Moore's paradox.

3.3. Lack of warrant

Another strategy that has been used to solve Moore's paradox involves the nature of warrant. Some claim that any Moorean judgment of something of the form of (O), (C), or (N) cannot be properly warranted.[111] You might argue that judgment of one of the conjuncts overrides or defeats any justification you might have for the other conjunct of the judgment, or you might argue that any justification you have for your judgment of one of the conjuncts also justifies you in denying the other conjunct.

This strategy also cannot succeed in providing a full solution to Moore's paradox. It also runs into a version of the attitude problem, and requires a solution to it to amend the full explanation of Moorean absurdity. It does so because there is no straightforward justificatory relationship between world-directed judgments (of *p* or *it's not the case that p*, for example) and self-attributions of belief (of *I believe that p*, or *I believe that it's not the case that p*, or *I don't believe that p*), as demonstrated in Chapter 2. There is also no straightforward justificatory relationship between the self-attributions themselves and the world-directed judgments in the other direction. There is a significant relationship between the fact that you *judge* the world-directed contents and your believing them, but that is precisely what we need a theory of self-knowledge to explain. Without that more substantive theory, this proposal cannot get off the ground.

This version of the attitude problem does not quite arise if you take the justificatory relationship to have a different structure. You might accept what Williams has called "Evans's principle": "Whatever justifies me in believing that *p* also justifies me in

[110] The attitude problem also arises for Heal's (1994) account, for Sorensen's (1988) account, and for Hintikka's (1962) account.

[111] See, e.g., de Almeida (2001), Williams (2004), and the discussion of both in Green and Williams (2007), pp.16-17.

believing that I believe that *p*."[112] Here the suggestion is not that one of the conjuncts of a Moorean absurdity might *itself* justify a judgment that comes into tension with the other conjunct. Rather, the suggestion is that further beliefs justify both the first conjunct of a Moorean absurdity and some content that directly or indirectly contradicts the second conjunct of the Moorean absurdity.

But the attitude problem arises once again in this context too. It is one thing to have justification for believing that you believe that *p*, and another to use it. How can the justification that you have for believing that *p* come into play in self-attributing a belief—in some way that would allow you to make a judgment that conflicts with the second conjunct of the Moorean absurdity? This is just the question of self-knowledge that I answered in the first two chapters of this dissertation.

Once again, this strategy used to solve Moore's paradox is impotent without an associated theory of how you can self-attribute beliefs knowledgeably.

3.4. Lessons

This survey of a few kinds of proposed solutions to Moore's paradox suggests that it is not easy to solve the paradox without a full theory of self-attribution of belief. Though it is possible to demonstrate that Moorean judgments violate various norms even without that further theory, the norm violations at issue are often not sufficient to explain actual absurdity in Moorean judgment.

It is not possible to survey all the proposed solutions to Moore's paradox here, but I hope that this discussion demonstrates why it is important to have a theory of self-knowledge in place to understand Moorean absurdities properly. The solution that I have proposed nicely ties together the facts that explain authoritative self-knowledge of belief with the facts that explain why it is absurd to judge or to assert something of the form of (O), (C), or (N). The facts that explain self-knowledge of belief are just a subset of the ten facts required to explain the absurdity involved in judging or asserting Moorean contents.

These facts can also be generalized in order to specify what would be required for some mental state *other* than belief to be transparently self-attributable in the same way as belief—and what would be required for there to be associated Moorean absurdities involving self-attributions of this other mental state as well. In the following section, I will generalize these facts in order to showcase this last advantage of the view I have proposed in this chapter.

4. Generalization

In Section 1 of this chapter I listed six facts that together sufficed to imply and explain why belief can be transparently self-attributed in a way that generates authoritative knowledge. Since we have characterizations of those facts, we can generalize them to determine whether any other mental states—or mental actions—can be self-attributed in similar 'transparent' ways, which also generate authoritative knowledge. We can also generalize the four additional facts used to solve Moore's paradox in Section 3 in order to see whether these other mental states or actions have associated Moorean absurdities.

[112] Williams (2004), p.348.

I'll start by generalizing the original six facts that sufficed for explaining transparent self-attribution of belief and its generation of authoritative knowledge.

4.1. The key facts in general form

Meeting the following six conditions suffices for some mental state or mental action *M* to be transparently self-attributable in a way that generates authoritative knowledge:

1. There is a kind of mental action *a* (with content *p*) the performance of which is sufficient for the contemporaneous existence of *M* with the same content, *p*.
2. You have control over your performance of this *kind* of mental action—that is, whether you are engaging in mental action of kind *a*. That implies that you can intentionally engage in *a*-ing. It also implies that when you do that, you know what you are doing *as a*-ing (or in equivalent terms).
3. The practical knowledge you have of what you are doing is already first-personal and immune to error through misidentification; it involves no positive identification of yourself among objects.[113]
4. To have the *M* concept, you must recognize that doing *that* kind of thing in thought—*a*-ing—is the appropriate kind of mental action to use to self-attribute states or actions of kind *M*.
5. The successful intentional performance of an *a* with content *p* involves consciousness with the content *p*.
6. Your mental action of *a*-ing with content *p* can, in the relevant intentional context, be one and the same event as that of your judging "I *M* with content *p*." That is, the mental action of *a*-ing with content *p* can be properly embedded into an ongoing mental task that is accomplished just by making this embedded action *a*.

Any mental state or action *M* that meets the conditions above is one that you can transparently self-attribute in a way that generates authoritative knowledge. That transparency method will work as follows:

- First, one consciously and intentionally sets out to self-attribute an *M*—perhaps an *M* with a content meeting some particular content criterion.
- Second, one consciously and intentionally sets out to perform some mental actions *a* (where, if one attempts to self-attribute an *M* meeting a particular content criterion, one's intention is to perform an action *a* that has a content meeting that same criterion).
- Third, one actually does go on to perform some such *a* (where, in the case of trying to self-attribute an *M* with a particular content, that *a* meets the content criterion in question).
- Fourth, one self-attributes an *M* with the same content as that of the action *a*. I call any such self-attribution a *transparent self-attribution of an M*.

I'll rehearse the epistemic explanation once more, in general terms, to vindicate the

[113] This is true of all practical knowledge, and so this condition is trivially fulfilled for any *M*.

proposal that any *M*s meeting those conditions are knowable in this way. The explanation has two parts: an explanation of truth, and an explanation of warrant. It will be crucial, as in the case of belief, that undergoing the method just described—which I will call the "transparency method" for any given *M*—involves embedded mental action: if one performs the transparency method, one self-attributes the *M* already in performing the mental action *a* whose content it shares.[114]

Why is a transparent self-attribution of an *M* true? Precisely because the first condition holds: the performance of an action *a* is sufficient for the contemporaneous presence of *M*, and in performing *a* in the context of the transparency method for *M* one already self-attributes an *M* (at the same time as one performs the relevant action *a*).

Why is a transparent self-attribution of an *M* warranted? This question splits into three parts. A *self*-attribution of an *M* is warranted when produced in this way because this method of self-attribution involves no error-prone identification of oneself; one is entitled to some such *self*-attribution due to one's immunity to error through misidentification. A self- attribution of an *M*—as opposed to any other kind of mental state or action—is warranted because one knows one is *a*-ing and, by virtue of having the *M* concept, one recognizes that *a*-ing is appropriate for self-attributing mental actions or states of kind *M*. Each of these two facts is again guaranteed by conditions listed above: the second condition guarantees practical knowledge in *a*-ing, and the third guarantees connective insight between the concept of an *M* and *a*-ing. Self-attribution of some *particular M* is warranted because in performing some such action *a* involves consciousness with the content of that *a*.

Note the role of embedded mental action in this context. The mental action *a* is the embedded mental action here; it is embedded in the ongoing task of trying to self-attribute mental actions or states of kind *M*. For a thinker with the concept *M*, who recognizes that doing *a* is sufficient for the contemporaneous presence of an *M* with the same content, embarking on the task of performing some such *a* with a view to self-attributing an *M* makes it the case that the *a* she does then perform *also is* a judged self-attribution of that *M*. Her intentionally *a*-ing as a way of self-attributing an *M* makes it the case that this mental action *a is also* a judgment that she has the corresponding *M*.

An *M*'s meeting these six conditions also contributes substantively to the formulation of potential Moorean absurdities involving self-attributions of *M*. In particular, if mental actions or states of kind *M* also meet these additional four conditions, they will have associated Moorean contents that are absurd to judge or to assert:

7. In conjunctive judgment, you consider the matters of all conjuncts at once.[115]
8. Directly contradictory *M*s are attributable to one agent only if those *M*s manifest in distinct rational contexts.
9. To have the *M* concept, you must understand that it makes no sense to attribute directly contradictory *M*s to one person in the same rational context.
10. We assume of one another that we make transparent self-attributions of *M* by default, and only use other methods when this method is unavailable.

[114] For a definition of embedded mental action and the conditions under which one performs embedded mental actions, see the previous chapter.
[115] This condition is trivially fulfilled for any *M*.

The question now, of course, is whether there *are* any such *M*s other than belief. I think there are at least three: judgment, decision (to act), and intention. I'll address judgment in the next subsection, and I'll address decision and intention in Chapter 4.[116]

4.2. Judgment

The extrapolation to the case of judgment is fairly simple, so I'll run through it here. Judgment meets the ten conditions listed above in the following ways:

1. There is a kind of mental action *a* (with content *p*) the performance of which is sufficient for the contemporaneous existence of *M* with the same content, *p*.

For judgment, this is trivially the mental action of judgment.

2. You have control over your performance of this *kind* of mental action—that is, whether you are engaging in mental action of kind *a*. That implies that you can intentionally engage in *a*-ing. It also implies that when you do that, you know what you are doing *as a*-ing (or in equivalent terms).

You have strong control over whether you are judging, as argued in Chapter 1.

3. The practical knowledge you have of what you are doing is already first-personal and immune to error through misidentification; it involves no positive identification of yourself among objects.

This condition is trivially fulfilled for any *M*.

4. To have the *M* concept, you must recognize that doing *that* kind of thing in thought—*a*-ing—is the appropriate kind of mental action to use to self-attribute states or actions of kind *M*.

To have the concept JUDGMENT, you must recognize that judging is appropriate to self-attribute judgment. This is straightforwardly necessary for having the concept.

5. The successful intentional performance of an *a* with content *p* involves consciousness with the content *p*.

As argued above, intentional judgment that *p* involves consciousness with content *p*.

6. Your mental action of *a*-ing with content *p* can, in the relevant intentional context, be one and the same event as that of your judging "I *M* with content *p*." That is, the mental action of *a*-ing with content *p* can be properly embedded into an ongoing mental task that is accomplished just by making this embedded action *a*.

[116] Supposition; Green (2007)

For the same reasons as with attribution of belief, the mental action of judging that *p* can also be a mental action of judging "I judge that *p*."

7. In conjunctive judgment, you consider the matters of all conjuncts at once.[117]

This is still true. This condition is trivially fulfilled for any *M*, including judgment.

8. Directly contradictory *M*s are attributable to one agent only if those *M*s manifest in distinct rational contexts.

One agent cannot judge that *p* and judge that it's not the case that *p* in exactly the same rational context. (Note here that what it takes for two judgments to be directly contradictory is just the same as what it takes for two beliefs to be directly contradictory.) Her evidence, her valuation of risk, her attention patterns, or something else relevant to her rational resources must change in order for her to change her mind.

9. To have the *M* concept, you must understand that it makes no sense to attribute directly contradictory *M*s to one person in the same rational context.

To have the concept JUDGMENT, you must understand that it makes no sense to attribute directly contradictory judgments to one person in the same rational context. That is, accepting condition 8 for judgment is part of having the concept JUDGMENT.

10. We assume of one another that we make transparent self-attributions of *M* by default, and only use other methods when this method is unavailable.

This is true for judgment as it is for belief.

The fact that judgment meets all ten of these conditions implies both that there is a transparent way of self-attributing judgments that grants authoritative knowledge, and that there are Moorean absurdities that involve judgment attributions instead of belief attributions. Both are fairly easy to extrapolate from the case of belief. A transparency method for judgment involves setting out intentionally to make judgments to self-attribute as such, rather than setting out intentionally to make judgments in order to self-attribute a belief as such. A judgment that *p*, in this context, can also be the same mental action as a self-attribution of that very judgment that *p*.[118]

The Moorean absurdities involving judgment attribution take the following omissive, commissive, and negative-commissive forms:

(O_J) *p*, but I don't judge that *p*.
(C_J) *p*, but I judge that it's not the case that *p*.
(N_J) it's not the case that *p*, but I judge that *p*.[119]

[117] This condition is trivially fulfilled for any *M*.
[118] Compare Burge (1996).
[119] Compare the same, with "think" in the place of "judge." "Think" can be ambiguous between "believe" and "judge" readings, but since there are three forms of Moorean absurdity for each of

The explanation of why it is absurd to judge or assert something with one of these forms proceeds in the same way (mutatis mutandis) as the explanation provided in Section 3 above. Part of the point of laying out the four extra conditions—on top of the original six key facts that support authoritative and knowledge self-attribution—on Moorean absurdity is to avoid the necessity of re-explaining the absurdity of judging and asserting various contents of Moorean forms with new attributions.

Note that it is generally the case that for any mental action of kind *a* that slots into the ten generalized conditions for attribution of a distinct mental state *M*, there should also be a transparent way of self-attributing *a* itself, and associated Moorean absurdities. This is just what we have seen for judgment. Judgment is a type of mental action *a* that slots into the ten conditions as belief meets them. judgment itself can be transparently self-attributed in a way that generates authoritative knowledge, and judgment itself has associated Moorean contents that are absurd to judge or to assert.

Conclusion

In this chapter I identified the six conditions that belief meets that together imply that you can transparently self-attribute beliefs to gain authoritative knowledge of what you believe. Since belief also meets four additional conditions, there are Moorean absurdities involving belief. The ten conditions together explain why it is absurd to judge or to assert any such Moorean content involving a belief attribution. The explanation that results is stronger than many other varieties of explanations that have been proposed as solutions to Moore's paradox.

After providing this solution to Moore's paradox, I generalized the relevant ten conditions and demonstrated how they apply to the case of judgment as well. Judgment can be transparently self-attributed in a way that yields authoritative knowledge. There are also Moorean absurdities that involve judgment attributions instead of belief attributions. In the next chapter, I tackle decision and intention.

---

belief and judgment, Moorean contents with "think" in the place of "judge" should be absurd to judge or to assert as well.

Chapter 4. Decision and Intention

In the Introduction I said it is important, when considering questions about how we self-attribute beliefs, to have a generality check on the view we endorse. In particular, it seems that the propositional attitude of intention has some relevant commonalities with belief. Intentions can be transparently self-attributed by way of decisions about what to do. We also have first-person authority about intention. By default, we take people's first-personal self-attributions of intention to be true, and we privilege these first-personal self-attributions over third-personal attributions of intention.

In this chapter I use the framework I have developed thus far to generalize my views about transparent self-knowledge and Moorean absurdities to the cases of intention and decision about what to do. In Section 1, I'll show how intention and decision each meet the six conditions that I set out earlier as necessary and sufficient for authoritative transparent self-knowledge. Then I'll use these facts to explain how you transparently self-attribute intentions and decisions. In Section 2, I'll compare the resulting view to other explanations of how we know what we intend to do in order to showcase the distinctive advantages of the view I have proposed. In Section 3, I will show that intention also meets the four additional conditions that helped explain Moorean absurdities for belief and judgment in the previous chapter. I'll consider candidate Moorean absurdities for intention. Though there are some conjunctive contents involving intention attributions that can be true even though they are absurd to judge, they are not, strictly speaking, Moorean absurdities for intention. That is because their absurdity cannot be explained in the same way as Moorean absurdities for belief. I conclude that the failure to find genuine Moorean absurdities for intention suggests quite strongly that intention is not a kind of belief.

1. Transparent self-knowledge

In the previous chapter I showed that meeting six general conditions is necessary and sufficient for some mental action or state *M* to be transparently self-attributed in a way that yields authoritative knowledge. The mental state I am currently considering is that of intention. In this section I show that intention and decision each meet the relevant six conditions. Then I'll explain how transparent self-attributions of either one can constitute authoritative self-knowledge of what the agent has decided, or what she intends to do.

1.1. The conditions on transparent self-knowledge

First, let's show how intention meets the relevant six conditions:

1. There is a kind of mental action *a* (with content *p*) the performance of which is sufficient for the contemporaneous existence of *M* with the same content, *p*.

For intention, this mental action is deciding what to do. Deciding to do something is sufficient for having a contemporaneous intention to do that (under the same description).

That is: deciding, at time *t*, to Φ is sufficient for having an intention, at time *t*, to Φ.[120]

2. You have control over your performance of this *kind* of mental action—that is, whether you are engaging in mental action of kind *a*. That implies that you can intentionally engage in *a*-ing. It also implies that when you do that, you know what you are doing *as a*-ing (or in equivalent terms).

You have strong control over whether or not you are engaged in deciding what to do at any time: merely *trying* to decide what to do ensures that you have thoughts with that attitudinal aspect. There is more to be said to argue for this claim; I say more below.

3. The practical knowledge you have of what you are doing is already first-personal and immune to error through misidentification; it involves no positive identification of yourself among objects.[121]

Practical knowledge is always like this, no matter what its object is. Practical knowledge is, by definition, non-observational. No observation is used to identify oneself to know of oneself that one is doing something. No other form of identification is used either. This condition holds for intention as it holds for any other mental state or action.

4. To have the *M* concept, you must recognize that doing *that* kind of thing in thought—*a*-ing—is the appropriate kind of mental action to use to self-attribute states or actions of kind *M*.

Here this means: to have the concept INTENTION, you must recognize that deciding what to do is the appropriate kind of mental action to use to self-attribute intentions. You could not, for example, recall facts you know, or simply imagine scenes, in order to self-attribute intentions. If you tried to do that, you would not have the concept INTENTION.

5. The successful intentional performance of an *a* with content *p* involves consciousness with the content *p*.

Successful decision to Φ, when you are intentionally engaged in deciding what to do, involves consciousness with the (partial) content Φ. Your decisions made unintentionally might not have this feature. If you make unconscious decisions, those decisions certainly will not have this feature either. But it is true that when you are *intentionally* deciding what to do, the decisions you make to fulfill your intention have conscious contents.

6. Your mental action of *a*-ing with content *p* can, in the relevant intentional context, be one and the same event as that of your judging "I *M* with content *p*." That is,

---

[120] It is more natural here to treat intentions as having contents that are actions, rather than propositions, although the conditions are written in terms of propositional contents. All the claims I make about the contents of intentions can be reframed in propositional terms—e.g. my intending to Φ is intending that I Φ (using the first-personal concept).

[121] This is true of all practical knowledge, and so this condition is trivially fulfilled for any *M*.

> the mental action of *a*-ing with content *p* can be properly embedded into an ongoing mental task that is accomplished just by making this embedded action *a*.

Here this means that deciding to Φ can, in the relevant intentional context, be one and the same event as your judging "I intend to Φ." This may be a little bit more difficult to see here, because a decision is one and the same action as a judgment, whereas in the prior case explained in Chapter 2, a judgment with one content was one and the same action as a judgment with another content. The two actions being identified on this theory are of two different kinds, whereas previously they were of the same kind. But that does not present any impediment to understanding the actions in question here. What is important is that the content of the embedded mental action be of the right kind to contribute to the content of the embedding action. Here, a decision to Φ is being used to determine the content of an intention one is self-attributing—an intention to Φ. That makes sense of the identification between the two actions in this intentional context.[122]

Most of these facts are fairly straightforward in their application to intention. The one that requires further argumentation is the fact about control: you have strong control over whether or not you are engaged in deciding what to do. Let's return to this point now.

We do not have on hand a tight and simple characterization of decision among other mental actions as we had on hand a tight and simple characterization of judgment among other mental actions.[123] I used that characterization to argue for the claim that you have strong control over whether or not you are judging when you are trying to judge as such.

However, we can rely on a rough and ready characterization of decision for the time being. To model this characterization on judgment, we should give the characterization both a normative and a descriptive component. Decisions are the mental actions that are normatively and descriptively guided by satisfiability by the agent. Decisions are objectionable or criticizable on the grounds that they are unsatisfiable—that the agent has decided to do something that she cannot do. An agent's decision to Φ cannot coexist with any conscious belief that it is impossible for her to Φ (in the circumstances in which she would intend to Φ, as specific or as vague as those are). This is a natural adaptation of the view that belief is normatively and descriptively guided by truth, as it reverses the 'direction of fit' of the relevant attitude in a way that respects that intentions represent things as *to be done*, whereas beliefs represents things as *already being so*.[124]

Given this characterization of decisions, we can also characterize intentions as the mental states that are normatively and descriptively guided by *the thing to do*.

This is unlikely to be the last word on the nature of decision (or intention). You can

---

[122] As noted in Chapter 2 for the case with two judgments being one and the same mental action: it is also important that the two actions be actions of one and the same agent. You cannot embed actions of your own in another's ongoing mental task, or vice versa.

[123] Shah (2008) actually offers a minimal characterization of intention: intention is the mental state that is "correct if and only if it is not the case that one ought not to perform the action that is its object" (p.12). I am not convinced by this characterization, partly because the idea of an intention being correct is underspecified.

[124] It is true that this characterization makes intentions less normatively constrained than beliefs, as (roughly put) more is possible than is true. This should not be a disadvantage. It seems indeed that intentions *are* less normatively constrained. This point will return later on in this paper when we consider Moorean absurdities for intention.

decide to do things without guiding your decision by what the thing to do is. You can decide to do something *just because*, or *because you feel like it*; you can even decide to do something in full knowledge that it is *not* the thing to do. These kinds of cases pose *prima facie* problems for this characterization of decision among mental actions.[125]

However, the details of the proposal need not concern us here. All that we need to recognize here is that there must be (de dicto) *some* normative and descriptive component to the characterization of decision. That much can be seen even without settling what these normative and descriptive components actually are.

First, consider the normative component. Decisions to act would not be what they are if they were not assessable by various standards, including standards of what is possible to do, and what is right for the agent to do, and what is in her best interests. These aren't standards that apply to all other mental actions too. Consider judgments; something can be *true* despite its being not right, or not best for the agent, etc. any mental acts that are not assessable by these standards are thereby not decisions to act at all. But these practical standards of assessment do need to apply to mental acts for them to qualify as decisions at all. We can be more sure that there *are* some such practical standards than we are confident in any particular characterization of those standards, e.g. in terms of what is best for the agent to do (as I have suggested above).

The descriptive component is more holistic but also necessary. Someone who was completely incompetent in taking into account various practical reasons, and possibilities for herself, and so forth, would not be identifiable as someone who was making *decisions* at all. Even if she were *trying* to hold herself to the standards that accord with the standards by which we assess decisions, the mental acts by which she does this would not count as decisions if their production by practical reasoning was entirely unconnected to the reasons she has to act, the facts of her situation, the question of what is best for her, and so on. This does not rule out mistakes of various kinds, but it's not completely superfluous as a condition on what decisions *are*.

As with judgment, what counts as enough competence in practical reasoning is a deep and difficult question. I cannot do justice to it here. But there is nonetheless a substantive point to be made about the level of competence required to be making decisions at *all*: the condition is at least as restrictive as that on the concept DECISION itself. What is good enough to qualify as descriptively guiding one's decisions according to the relevant standards is no stronger than what is good enough to have the concept DECISION.

The proposal I am making here is that there is *some* correct individuation of decision among mental acts of the following form: decision is the mental act that is normatively and descriptively guided by *X*. For our purposes, I have proposed that we can consider a particular value for *X* here: the thing to do. But I am less committed to this particular value for *X* than I am to the general shape of the individuation of decision.

Once we know that a full characterization of decision must have both normative and descriptive components, we can then turn to argue that we have strong control in deciding what to do (as opposed to engaging in other kinds of occurrent thought).

Consider someone who intentionally sets out to decide what to do. Anyone who can

[125] These prima facie problems are not immediately fatal to the proposal. They can be accommodated, for instance, by saying that even those decisions that are not themselves decisions to do *the thing* to do are nonetheless *guided* by the thing to do—and assessable according to this standard. Compare Gibbard (2003) on the thing to do.

do that has the concept DECISION. If she tries to decide what to do (as opposed to engaging in other kinds of thought), does that guarantee that she does engage in the kind of thought she tries to engage in—i.e. deciding what to do?

To show that she does, we must show that she meets the two conditions on decision outlined above. There is a normative and descriptive component to the individuation of decision. Just by holding herself to the standard that is specified as the normative standard that individuates decision, she guarantees that the thoughts performed with her aim are assessable by that standard. The descriptive condition—that she be competent enough in weighing reasons and guiding her thought in practical reasoning—is guaranteed to be met because she has the concept DECISION. Above, we saw that the level of competence required to be able to count as making decisions when you try to do so is no stronger than the level of competence required to have the concept DECISION. That rules out as impossible the error you might make by trying to make decisions as such but doing such a hopeless job that you do not even qualify as making decisions at all (let alone making *good* or *poor* decisions). That means that anyone who tries to engage in decision as such in occurrent thought is guaranteed to be engaging in that type of thought. That is, each of us has strong control in deciding what to do as such.

Having strong control over whether you are deciding what to do does not imply any kind of voluntarism about deciding what to do. You cannot arbitrarily choose, for some Φ, that you will decide to Φ. A 'decision' made like that would not be a decision at all, as it would not be normatively and descriptively guided in the way decisions are necessarily guided. Even though you can intentionally engage in deciding what to do, and you control *whether* you are deciding what to do, you do not control *what* you decide to do.

Thus far in this section I have argued that intention meets the six general conditions that I set out in the previous chapter. It is also true, for almost identical reasons, that decision itself meets the six conditions as well.

What this implies is that both intentions as such and decisions as such should be transparently self-attributable, and that those transparent attributions of intention or decision should constitute authoritative self-knowledge for the agent.

1.2. Transparent self-attribution

How can you transparently self-attribute an intention to Φ, or a decision to Φ? In this subsection I will explain each how you can make either kind of transparent self-attribution, and why those transparent self-attributions constitute authoritative self-knowledge of what you intend to do, or what you decide to do.

Here is what you can do to self-attribute an intention:

- First, you consciously and intentionally set out to self-attribute an intention, perhaps meeting some content criterion (e.g. an intention to go to the party or not).
- Second, you consciously and intentionally sets out to make a decision meeting the same content criterion (e.g. a decision to go to the party or not).
- Third, you actually do go on to make a decision one way or the other (e.g. a decision to go to the party).
- Fourth, you self-attribute an intention with the same content as the decision (here, an intention to go to the party).

Recall the importance of embedded mental action here. In this context, one performs the third and fourth 'steps' all at once. Due to your setting out to make a decision whether to go the party *in order to* self-attribute an intention on the matter, your decision to go to the party *is also* a self-attribution of an intention to go to the party.

Why does a transparent self-attribution of an intention—that is, any self-attribution of intention made in this way—constitute authoritative knowledge? In order to explain that, we can appeal to the same kinds of facts that were relevant in explaining why transparent self-attribution of belief constitutes authoritative knowledge of what you believe.

To see why any transparent self-attribution of intention must be *true*—at least at the moment at which it is made—recall that decision to Φ is sufficient for contemporaneous intention to Φ.

There are several aspects of warrant that you have for any transparent self-attribution of intention. You are entitled to your *self*-attribution of intention here because your method of attribution is immune to error through misidentification. You are warranted in self-attributing an *intention* rather than any other mental state due to your practical knowledge of your deciding what to do, and your conceptual entitlement that connects that activity with the concept INTENTION. You are warranted in self-attributing a particular intention to *Φ* because your decision, as an intentional decision, has a conscious content. The warrant you have for your overall transparent self-attribution of intention to Φ cannot be Gettierized in any way. So long as your self-attribution is transparent in the way defined above, there is no question of your having the *wrong* sort of warrant, or deriving your self-attribution from false lemmas, and so on.

The authority that attaches to this kind of self-attribution of intention is also explained with reference to the same kinds of facts that were relevant in the case of belief. Your self-attributions of intention are presumed to be true because they must be, given the sufficiency of decision for contemporaneous intention. They are accepted over others' attributions of intention to you because nobody else has such a secure way of attributing intentions to you. Nobody else can perform any mental action of any kind that is sufficient for your intending to Φ at the same time, and nobody else can initiate *your* mental actions. The practical knowledge you have of what you are doing, and the consciousness of the content of the decision you make, are special to the first-personal perspective you have on your own intentional mental actions (here, decisions).

These facts, again, would only vindicate the presumptions involved in first-person authority if we accept a further important fact: we presume of each other that we make *transparent* self-attributions of intention by default. If we could not assume this of one another, we could not use the epistemic merits of the transparency method to rationalize the deference we give to first-personal self-attributions of intention in general.

The fact that our explanation once again exploits important facts about what it *is* for some thought or action to be yours, as opposed to someone else's, is a good sign that we have gotten onto the right explanation of first-person authority in this case.

There is one last important fact to note here. The view I am proposing implies that practical knowledge—knowledge of what you are doing, when you are doing it intentionally—is more basic than knowledge of intention itself. This rules out views of practical knowledge on which prior knowledge of intention is required for knowing what you are doing when you are doing something intentionally. I think it is correct to rule

these views out. Nonetheless, it is important to note this substantive consequence of the position I have advanced here.

All of the same points above apply *mutatis mutandis* to the simpler case of decision. You can transparently self-attribute decisions in almost exactly the same way in which you self-attribute intentions. Transparent self-attributions of decision also constitute authoritative knowledge.

2. Comparison to other views

In the last section I explained how you can transparently self-attribute intentions. When you do that, your transparent self-attribution of an intention constitutes authoritative self-knowledge. In this section, I'll compare the view I explained in the last section with some other prominent proposals about self-knowledge of intention. I address views proposed by Richard Moran, Alex Byrne, Sarah K. Paul, and Kieran Setiya.[126]

Here I assume, for the purposes of assessing these views, that it is true that we usually know what we intend to do, and we know that with first-person authority. Some philosophers, including Jay Wallace, have suggested that it is actually impossible to have an intention you do not know you have, and that it is similarly impossible to think you have an intention that you do not in fact have.[127] I will not go so far as to make this claim in this chapter. One reason to worry about it is that it may be possible for agents to have intentions without having the concept INTENTION. If there could be agents like that, then it would be possible for an agent not to know her intentions as such. But not much will hang on this point, since we will mainly be concerned with the question of how agents with the concept INTENTION know what they intend in ordinary cases.

I will argue that each of the views I address below fails to solve a close cousin of the attitude problem that I discussed in Chapter 1. The new problem is just **the attitude problem for intention**. Each view assumes, rather than explaining, how one comes to realize that one *intends* to do something, rather than merely entertaining the thought of doing it, or thinking you *will* do something, but not intentionally. An explanation of self-knowledge of intention needs to explain how an agent distinguishes between states with these distinct attitudinal aspects, and the views I survey below do not explain how an agent can do that to self-attribute an *intention* as such.

I'll start by addressing what Moran has said about self-knowledge of intention.

2.1. Moran

As in his explanation of self-knowledge of belief, Moran takes it that your epistemic authority on the matter of what you intend is explained by your deliberative authority over what you intend.[128] You make up your mind about *what to do*. You deliberate about what to do in order to form and reaffirm your decisions in conscious thought. And in ordinary cases, such practical deliberation really does determine what you intend to do.[129]

For the same reasons as with belief, Moran's points about the responsibility you bear

[126] See Moran (2001), Byrne (2011), Paul (2012, 2015), Setiya (2011).
[127] See Wallace (2001), p.22.
[128] Moran (2001).
[129] This paragraph is adapted from that about Moran's view concerning belief in Chapter 2.

for your intentions, and the fact that you make up your mind in forming intentions, do not properly amount to a full explanation of how you know what you intend to do. You can be responsible for the way something is without knowing how it is. You can be the producer of something without knowing how it ends up in full detail.

Moran later amended his view to include an aspect of entitlement.[130] You are entitled to take it that your deliberations about *what to do* constitute your *intentions* to do things, he wrote, and in the ordinary case this really is true. But making use of this entitlement already requires a certain sensitivity to the attitudinal context of your practical deliberation. You must in some sense be sensitive to the fact that you are *deciding what to do*, rather than idly entertaining things you *could* do, in order to recognize that this is a context in which your deliberations determine what you intend. Since Moran has not explained this sensitivity itself, he has not solved the attitude problem for intention.[131]

2.2. Byrne and Setiya

Byrne offers an adaptation of his view about transparent self-attribution of belief as an explanation of self-knowledge of intention.[132] In order to self-attribute intentions in a knowledgeable and authoritative way, you can follow what he calls "the bouletic schema" in your reasoning: you can reason from "I will Φ" to "I intend to Φ." Performing this routine in moving from the first belief to the second does not constitute an *inference*, but it will almost always yield knowledge about what you intend to do.[133]

Byrne recognizes an immediate problem with this simplest characterization of the bouletic schema: you can judge of yourself that you will Φ when you do not in fact intend to Φ. You might judge, for example, that you will fail the exam, even if you do not intend to fail the exam. This is a serious problem for the bouletic schema as knowledge-conducive; a large class of beliefs about what you will do in the future do not constitute expressions of intentions.

At this point, Byrne's view faces a dilemma. On either horn of the dilemma, the view fails to solve the attitude problem—the same problem that faced Moran's explanation of how we know what we intend to do. The horns of the dilemma derive from the two proposals Byrne makes in order to amend the characterization of the bouletic schema.

The first proposal is just that the first belief in the routine should not be "I will Φ" simpliciter, but rather "I will Φ intentionally." Byrne quickly rejects this proposal, as he should. This proposal fails to solve the attitude problem by simply ignoring it. To believe

---

[130] Moran (2003, 2004).

[131] It is worth noting that I do not agree with Paul's (2012) criticisms of Moran's view. She claims that intention is not transparent in the way that Moran claims it is. But she supports that point by showing that what you do intend cannot be answered by asking what you have reason to do (p.334), what you ought to do (p.334), or what it would be best to do (p.335). You can intend to do things that do not meet any of these descriptions. That is true. But the best version of Moran's view takes the question of what you intend to do to be transparent to the question of what *to do*, which matter is not separable in the same way from what you intend to do. See Section 2.3.

[132] Byrne (2011, p.214ff).

[133] The bouletic schema, Byrne (2011) writes, is "*practically* strongly self-verifying: for the most part, if one reasons in accord with the schema (and is mindful of defeating conditions, for instance the one just noted), then one will arrive at a true belief about one's intention" (p.219).

that you will Φ intentionally is already to have a belief about your intentions. It builds in, rather than explains, self-attribution of a particular kind of attitude—here, intention.[134]

The other option, the one that Byrne accepts, is to restrict performance of the bouletic routine to some subset of cases that really are cases in which the Φ of "I will Φ" is something the agent intends to do, rather than merely *expects* that she will do, either intentionally or not. Those conditions are ones in which the agent does not believe "I will Φ" based on *evidence*, but rather believes it non-observationally.[135] Byrne reasons that the capable agent who is to gain any knowledge from performance of the bouletic routine must be someone who does *not* perform it in any contexts in which she believes the 'premise' ("I will Φ") based on evidence. As he puts it, the condition in which "I will Φ" is based on evidence is a *defeating condition* on the knowledge you might otherwise gain by use of the bouletic routine. You can still gain knowledge of what you intend by performing the bouletic routine, but only if you are sensitive enough to the defeating condition *never* to perform it when you have evidence for your belief "I will Φ."

There is a basic problem with the mechanics of this proposal, and there is a serious problem that is fatal to the proposal. The second is a version of the attitude problem.

Here is the basic problem with the mechanics of the proposal. You ordinarily *do* have evidence (and often good evidence) that you will do something that you also do intend to do. You can believe you will do something *both* because you intend to do it *and* because you have good evidence to believe you will do it. That means that in many cases the belief that begins the bouletic routine, I will Φ, may be based on good evidence that you will Φ as well as expressive of (or partly constitutive of) your intention to Φ. Figuring out whether your belief that you will Φ is based on that good evidence you have or not is less straightforward than Byrne makes it out to be.

The more important problem, however, is a version of the attitude problem that arises once again at this point. Realizing that a belief of yours is based on *no* evidence—as one, in some way, needs to do in order to use the bouletic routine—can lead to several distinct outcomes: giving up the belief, and (if Byrne is to be believed) attributing an intention to yourself. But what delineates those circumstances in which one gives up the belief because of lack of good evidence from those in which one goes on to self-attribute an intention? This has gone unexplained. It assumes the same sensitivity that it aims to explain—the sensitivity to whether your belief is expressive of an intention or not. Your capacity to perform the bouletic routine when and only when you really *intend* to Φ, rather than when you simply realize you have no reason for your preexisting belief that you will Φ, is just a capacity to track your own intentions as such. That is, once again, to ignore the attitude problem, rather than to solve it.

Both amendments to Byrne's proposal fail to solve the attitude problem. On either horn of the dilemma, we are not given a full explanation of how you know what you *intend* as such.

Setiya's position is very similar to Byrne's view, although he recognizes some key points about the nature of transparent self-knowledge that Byrne misses.[136] He recognizes

[134] In fact, Byrne (2011) rejects this proposal for less compelling reasons than this one (p.217).
[135] Byrne (2011) interprets "knowledge not based on observation" as "knowledge not resting on evidence" (p.218). This seems unobjectionable in this context.
[136] Setiya (2011).

that it is epistemically groundless: "it does not rest on quasi-perceptual appearances … or on inference from evidence of other kinds."[137] He also recognizes that you need not move *from* some judgment about what you will do *to* a further judgment about what you intend. The two facts can be recognized in one and the same thought.[138] His claim is that "the capacity to act for reasons, as a capacity to know what I am doing by intending to do it, is used to gain knowledge of my intention along with the belief about what is happening that this intention provides."[139]

The attitude problem arises for Setiya's view in just the same way as it arises for Byrne's view. When you know that you are doing something, you might know that simply by doing it intentionally (and having sufficient control over your success), or you might know it by observation. In order to complete his proposal, Setiya needs to explain how you can distinguish the cases. In order to solve the attitude problem rather than ignore it, this kind of view needs to do more than stipulate this basic awareness of the attitudes on which you act.

2.3. Paul

Paul agrees that intention can be transparently self-attributed by way of deciding what to do. She agrees with this point although she thinks that deciding to Φ is not sufficient for having a contemporaneous intention to Φ. However, it is true that "deciding to Φ is a way of forming an intention to Φ." The connection is robust enough, she thinks, to support this transparent self-attribution despite its not being "failsafe."[140]

However, Paul's view fails to solve the attitude problem by default. Paul simply *defines* decisions as "discrete, conscious mental events." As such, she says, they "are normally known to the thinker at the time of making the decision; if one does not know whether one has decided, one generally has not."[141] Here Paul conflates the consciousness of a decision—which is consciousness with the *content* of the decision—with consciousness of the attitudinal aspect of decision. One can make decisions, and even conscious decisions, without being aware that they are *decisions*, as opposed to mere entertainings of things one might do.

Paul's view might be patched up with recognition of some decisions as intentional mental actions over which the agent has control. If Paul recognized the importance of intentional mental action here, and so recognized the importance of practical knowledge here as well, she might solve the attitude problem in the same way I have done above. However, Paul decidedly resists any neat normative characterization of the nature of decision or intention.[142] Without seeing that normative constraints partly individuate decisions among mental acts and intentions among mental states, it is difficult to see why we have strong control over our engaging in *deciding what to do*, and thus difficult to see

[137] Setiya (2011), p.178.
[138] Setiya (2011) also recognized this point elsewhere for the case of belief. He has not, however, explained how such content plurality can be effected; he did not recognize the importance of intentional mental action here.
[139] Setiya (2011), p.194.
[140] Paul (2012), p.337.
[141] Paul (2012), p.338.
[142] Paul (2012), pp.332-5.

how we have practical knowledge in intentional decisions.

There is much more in Paul's discussion that agrees with the explanation I have given above. She accepts at least one normative condition on decision: "a thinker cannot decide to Φ if he does not believe it to be reasonably possible for him to Φ."[143] She also recognizes the importance of a conceptual connection between decision and intention: "the fact that making a decision to Φ is a way of forming an intention to Φ is contained in the very concepts of decision and intention."[144] To say that much is to approach the rough and ready characterization of decision I gave above. Despite these important realizations, though, Paul's discussion fails to *explain*, rather than *assume*, a crucial point: how you know, when you are deciding to Φ, that it is *deciding* you are doing. Paul's view also fails to solve the attitude problem for intention.

2.4. Lessons

Some of the disadvantages of the views discussed here come from failure to recognize important facts about the relationship between decision and intention. Decision to Φ at any one moment is sufficient for contemporaneous intention to Φ—even if that intention is abandoned later on. We saw that other kinds of judgments—e.g. the judgment that you will Φ—are not sufficient for having a contemporaneous intention. There is also another key fact absent from almost all the explanations of self-knowledge discussed above: we presume of one another that we make transparent self-attributions of intention by default. Without making this assumption of one another, we could not properly explain the interpersonal deference involved in first-person authority.

However, it is most important to the explanation of authoritative self-knowledge of intention that you can engage in *intentional* mental action, with control over the kind of thought you are engaging in, so that you have practical knowledge that captures the attitudinal aspect of your thought. Without recognizing the importance of intentional mental action and practical knowledge here, we could not solve the attitude problem for intention. The views offered by Moran, Byrne, Setiya, and Paul above all fail to solve the attitude problems for slightly different reasons. Any one of their views would benefit from recognition of the importance of intentional mental action.

3. Moorean absurdities

I have now explained how you can transparently self-attribute intentions, and I have explained why transparent self-attributions of intentions constitute authoritative knowledge. In this section, I find Moorean absurdities involving intention.

In Chapter 3, I showed that there are six necessary and sufficient conditions that some mental state or act must meet in order to be transparently self-attributed in a way that yields authoritative knowledge. I also showed that there are four further conditions of interest which will generate Moorean absurdities for any mental state or act that meets them. Intention and decision both meet these further conditions as well. The further conditions are these:

[143] Paul (2012), p.336.
[144] Paul (2012), p.339.

7. In conjunctive judgment, you consider the matters of all conjuncts at once.

All mental states and acts meet this condition trivially.

8. Directly contradictory *M*s are attributable to one agent only if those *M*s manifest in distinct rational contexts.

An intention to Φ and an intention *not* to Φ are directly contradictory intentions. These are, indeed, only attributable to one and the same agent if those intentions manifest in distinct rational contexts. The same point holds for decisions.

9. To have the *M* concept, you must understand that it makes no sense to attribute directly contradictory *M*s to one person in the same rational context.

This builds the previous condition into the concepts of INTENTION and DECISION.

10. We assume of one another that we make transparent self-attributions of *M* by default, and only use other methods when this method is unavailable.

This condition already had to be met in order to explain first-person authority about self-attributions of intention and decision.

In the case of belief, meeting these extra conditions let us explain what was absurd about judging or asserting Moorean absurdities for belief. Judgment also met these additional conditions, and we found Moorean absurdities involving judgment attributions. This suggests that there should be Moorean absurdities for intention as well.

How are we to go about looking for Moorean absurdities for intention? Recall that traditional Moorean absurdities have two forms, omissive and commissive:

(O) *p*, but I don't believe that *p*.
(C) *p*, but I believe that it's not the case that *p*.

Substitute some sentence in for *p* in either schema, and you'll get a sentence that can be true although it is absurd to assert or judge—but only in the first-person present tense.

It's not a good idea just to replace "believe" with "intend" in schemas (O) and (C) above. That is a way of producing sentences that are like Moorean absurdities for belief in some sense, but it is not a way to produce absurd sentences at all. *p* can be the case when you don't intend to bring it about; and *p* can be the case when you intend to make it that it's not the case that *p*.

A few distinct constraints will guide a more sophisticated search for Moorean absurdities for intention. In what follows, I will look for *sentence schemas* involving variables that range over verbs ("Φ"), just as (O) and (C) above involve variables that range over sentences ("*p*").[145] I am looking for sentence schemas, rather than individual

[145] It may be objected at this point that we need to treat intention as a *propositional* attitude in order to think about Moorean absurdities for intention. I do not think that considering examples involving variables over verbs, rather than variables over sentences, rules out that intention is a propositional attitude. It is often more natural to frame intention attributions with the locution

sentences, because there is some generality to the phenomenon of Moorean absurdities for belief: there are many values for *p* that would make an assertion of an instance of (O) or (C) absurd. That does not mean that perfect generality is required.[146] As we will see below, there are some cases in which restricting the range of a variable will uncover interesting candidates.

The second constraint on our search has to do with the forms of the schemas we seek. It is important to the original phenomenon of Moorean absurdities for belief that they come in several forms. In fact, Moorean absurdities for belief seem to me to come not only in the two forms listed above but also in another form:

(N) It's not the case that *p*, but I believe that *p*.[147]

We then have three original forms for Moorean absurdities for belief: omissive, commissive, and *negative-commissive*. Here they are all together:

(O) *p*, but I don't believe that *p*.
(C) *p*, but I believe that it's not the case that *p*.
(N) It's not the case that *p*, but I believe that *p*.

Each of these forms is essentially conjunctive.

The omissive form essentially involves one conjunct that expresses or implies the presence of a particular belief (that *p*) without mentioning intention, and another conjunct that consists in the negation of a self-attribution of the same belief (that *p*). Importantly, the negation in the second conjunct of the omissive form is outside the scope of the belief operator.

The commissive form essentially involves one conjunct that expresses or implies the presence of a particular belief (that *p*) without mentioning belief, and another conjunct that consists in a self-attribution of the negation of the same belief (that it's not the case that *p*). Importantly, the negation in the second conjunct of the commissive form is inside the scope of the belief operator.

The negative-commissive form essentially involves one conjunct that consists in the negation of what would be an expression or implication of a particular belief (that *p*) without mentioning belief, and another conjunct that consists in a self-attribution of the

---

"intend to Φ" instead of "intend that *p*" or "intend to make it the case that Φ," and naturalness is important here: we hope to be able to assess the absurdity of asserting various sentences, or judging their contents, without significant further theoretical commitment.

[146] As mentioned in a footnote above, 'perfect' generality is already ruled out for Moorean absurdities of the form of (O) and (C), since we must restrict the range of "*p*" in order to target the intended phenomenon—one that involves satisfiable truth conditions.

[147] You might take instances of (N) to be instances of (C) insofar as believing that *p* is tantamount to believing that it's not the case that it's not the case that *p*. However, given the opaque context of belief attributions, and the fact that not every value for *p* is equivalent to its double negation even when *p* has satisfiable truth conditions (as in cases of vagueness or indeterminacy), I will consider (N) as a separate form. This will matter to the following discussion, since there are cases in which closely related commissive and negative-commissive versions of candidate Moorean absurdities for intention differ in absurdity.

same belief (that *p*).

Specifying these forms helps to highlight a few key ingredients of Moorean absurdities for belief. These ingredients are: a statement that expresses or implies the presence of an attitude without mentioning that attitude; a conjunction; and a statement that involves (or negates) a self-attribution of belief. We can adapt these ingredients to put together a recipe for Moorean absurdities for intention. In order to find some *candidate* Moorean absurdity for intention, we need to identify a kind of statement that can express or imply the presence of an intention without mentioning intention, and then conjoin some such statement (or a negation of the same) with a self-attribution of the intention expressed or implied (or a negation of the same, inside or outside the scope of the intention operator, depending on the form).

As will become clear, most of the room for creative maneuver in the search that follows lies in the task of finding statements that express or imply the presence of an intention without mentioning intention. In one case, we will consider an expression that seems to express or imply the lack of an intention, rather than its presence.

What I have said so far specifies a recipe for candidate Moorean absurdities for intention, but it does not yet give us any guidance on how to assess some such candidate Moorean absurdity for intention once we have used the recipe to produce one. To assess the candidates once we have them on the table, I will use the following three fundamental features of Moorean absurdities for belief, also discussed above:

> *I. Satisfiability.* Moorean absurdities can be true; in other words, their contents have satisfiable truth conditions.
> *II. Absurdity in assertion and judgment.* Assertions and judgments of Moorean absurdities are absurd, but non-assertive (or embedded) utterances—and non-judgmental (or embedded) thoughts—with the same contents are not absurd.[148]
> *III. First-personal present tense.* It's not absurd to assert or judge any version of a Moorean absurdity that is not in the first person or not in present tense.[149]

Using these constraints, we will find several serious candidates for Moorean absurdities for intention. That is: we will find several conjunctive contents that seem absurd to judge, and thus also absurd to assert, even though they are satisfiable. The absurdity in question is constrained to the first-personal, present-tense contents as they stand.

However, I will argue in Section 3.6 below that these absurdities are not *Moorean* in the sense that is most relevant to our discussion. Even though important facts about intention explain the absurdity involved in asserting or judging any one of these things, we cannot explain the absurdity in the same way as we did in the case of belief.

The key fact that explains the asymmetry between Moorean absurdities for belief, and these other absurdities concerning intention, is that there is no judgment one can make that is tantamount to deciding to Φ. That suggests that intention is not a form of belief.

I'll begin by looking at one natural proposal about Moorean absurdities for intention.

---

[148] For example, it's not absurd to utter something of the following form: "Suppose the following. *p*, but I don't believe that *p*." Correspondingly, it's not absurd to suppose that much.

[149] It's not absurd to assert or judge any of these: "*p*, but you don't believe that *p*," or "*p*, but she believes that it's not the case that *p*," "*p*, but I didn't believe that *p*," "*p*, but I will believe that it's not the case that *p*," etc. The plural first-personal version is a little odd; cf. Sorensen (1988).

3.1. I will Φ, I am Φ-ing, I shall Φ

A common thought, famously expressed by Anscombe (1957), is that the paradigmatic expression of an intention is a future-tense statement about what one *will* do.[150] If that kind of statement can express intention, then we should try conjoining it with self-attributions of intentions (or negations of the same) to produce candidate Moorean absurdities for intention. Consider, then, the following adaptations of (O) – (N) above:

(1O) I will Φ, but I don't intend to Φ.
(1C) I will Φ, but I intend not to Φ.
(1N) I will not Φ, but I intend to Φ.[151]

Instances of each of these schemas can have satisfiable truth-conditions. Consider the following instances, using the verb "to die" as a value for Φ:

(1O.i) I will die, but I don't intend to die.
(1C.i) I will die, but I intend not to die.
(1N.i) I will not die, but I intend to die.

Each of these sentences can be true. (1O.i) can be true because you can die unintentionally. (1C.i) is true if dying is in your future, but you try to avoid it—or you aren't aware that it is inevitable, and so you make an effort to stay alive forever. (1N.i) can be true because you can fail to do something you intend to do.

Are these sentences absurd to assert or to judge? Certainly (1O.i) is not absurd at all. From your own perspective, you can understand that you will do something that you do not do intentionally. Dying is a paradigm case: many people die without intending to die.

Asserting either (1C.i) or (1N.i) is a little stranger. In making an assertion of either one, you represent yourself as thinking that your intention will be thwarted; in judging either, you think that your intention will be thwarted. On the model of Moorean absurdities for belief, we might say that the first conjunct in each *expresses* or *implies the presence of* a belief that you will die, in the commissive case, or a belief that you will not die, in the negative-commissive case.

Is it possible to think you will die, while still intending not to die? Is it possible to think that you will *not* die, while still intending to die? This is a point of potential controversy, and it matters a lot here. This point might decide whether it is just strange, or fully *absurd*, to assert (1C.i) or (1N.i). If it is fully *impossible* to think, in this clear-eyed way, that some particular intention of yours (de re) will fail, then it would seem that asserting or judging either (1C.i) or (1N.i) would be absurd. If so, (1C) and (1N) might be

[150] The canonical expression of this point in Anscombe (1957) actually involves the formulation "I am going to Φ." Nothing in the following discussion will rest on the fact that I use "I will" instead of "I am going to"; all the same points apply, since "I am going to"—like "I will"—can also be used to express 'pure' prediction rather than intention (e.g. "I am going to fail").

[151] As before: (1O), (1C), and (1N) are sentence schemas, and Φ is a variable ranging over English-language descriptions of actions (understood quite broadly, as will become clear below). For the sake of discussion, though, I exclude as possible values for Φ all descriptions of patently impossible or self-defeating actions (e.g. "do something impossible" or "die and not die").

forms of Moorean absurdities for intention although instances of (1O) are not absurd.

Let's think about considerations that bear on this question. I cannot review them all here, but I will touch on a few ways to argue for either side.

There may be good reason to think that that is impossible to think you will do something (*de re*) and retain the intention not to do it (and likewise with switched negations). Having an intention, you might say, involves seeing the future as fundamentally open for you to shape; and if you think what you intend to bring about will not happen, you do not see the future as open in this important way.

Perhaps it is not *impossible* to have a belief that you will die and an intention not to die, because sometimes your beliefs and intentions don't come into 'direct contact.' They may be in some sense isolated from each other, and they may only come up for you consciously in distinct contexts. That may be true, but perhaps it is not possible to have some such belief and some such intention once you bring them in contact with one another in the same context. If you had that kind of view, then judging (1C.i) or (1N.i) would still be absurd—because judging that would involve bringing together a belief and an intention that cannot coexist once considered together. Asserting the same would then imply that you brought together attitudes that cannot coexist together in thought.

On the other hand, if these attitudes can coexist together in one and the same thought, then it is not necessary to admit absurdity in judging and in asserting (1C.i) or (1N.i). There are likely many different ways of holding this view in a reasonable way; let's review two of them.

Here is one way of explaining how a belief that you will do something, and the intention not to do it, can coexist in thought. You might think that it is sufficient to believe that you will do something that you have more confidence in the proposition that you will do it than you have in the proposition that you will not do it. You might also think that you do not need to have more confidence in the proposition that you will succeed in some endeavor than the proposition that you will fail at the same endeavor in order to *intend* to succeed in that endeavor. If you think both of those things, then perhaps you can bring together the belief that you will die and the intention not to die in the same context without absurdity.

If so, then it is also possible to judge or assert (1C.i) or (1N.i) without absurdity. Saying or thinking something like (1C.i) or (1N.i) is a way of manifesting an expectation that your intention will likely be thwarted. To see this interpretation more clearly, try flipping the conjuncts of (1C.i): asserting "I intend not to die, but I will die!" may seem more natural as an expression of an expectation of failure. The same goes for (1N.i): consider asserting "I intend to die, but I will not die." You might assert something like (1C.i) or (1N.i) if your expectation that you will fail is based on your assessment of some external conditions—it's a long shot, you might say—or on an assessment that your own willpower will falter. Just how much confidence in your success you need to have in order to have a genuine intention is, as mentioned above, a matter of significant philosophical controversy.[152]

Here is a second way of bearing out the contention that it is not absurd to assert or to judge (1C.i) or (1N.i). You could anchor an understanding of what is said in an assertion of (1C.i)—or what is thought in a judgment of (1C.i)—with reference to a closely related content: "I will die, but I am trying not to." Asserting or judging that seems much less

[152] See the literature on lotteries and slim chances.

strange than asserting or judging (1C.i). However, you might argue that trying to Φ manifestly involves having an intention to Φ. It doesn't seem that you can *try* to do anything without having *some* intention—and what else would your intention be, other than the intention to Φ? If you argued in this way, you might take it to be unacceptable to give distinct verdicts of absurdity for "I will die, but I am trying not to," and (1C.i): "I will die, but I intend not to die." If you did that, and you took the verdict of non-absurdity in the former case to involve a stronger, or more reasonable, intuition, then you could conclude that it is not really absurd to assert, or to judge, (1C.i). The marked quality of (1C.i) might come from 'mere' linguistic convention in the way we use the words 'try' and 'intend.' You might make an exactly analogous argument for (1N.i).

None of these points are uncontroversial, and the arguments I have sketched here are not ultimately decisive on these issues. Suffice it to say, for present purposes, that there are reasonable ways of making out assertions and judgments of (1C.i) and (1N.i) *not* to be absurd, even though there are also reasonable ways of making them out to be genuinely absurd as well.

Does that mean that we should give up on the forms (1O), (1C), and (1N) as forms of Moorean absurdities for intention? Not quite. There may be ways of further constraining these forms that would reveal more interesting candidates to consider. Let's try to diagnose what makes (1O.i), (1C.i), and (1N.i) controversial as examples of sentences that are *absurd*, rather than just marked or odd, to assert or judge.

One important factor here is that we can interpret the future dying mentioned in the first conjuncts of (1O.i) – (1N.i) as non-intentional. In fact, the second conjuncts of each of these sentences seem to *force* the interpretation of that dying as unintentional or counter-intentional—that is, contrary to one's actual intentions, a matter of failure.[153] This looks like it may pose a general challenge to our project of finding genuine Moorean absurdities for intention: the **problem of forced interpretation**.

Let's try to avoid this problem with verbs I'll call "Anscombe verbs": verbs that denote constitutively intentional actions.[154] One such verb is "to marry." You constitutively cannot marry without intending to marry. Consider, then, the following:

(1O.ii) I will marry, but I don't intend to marry.
(1C.ii) I will marry, but I intend not to marry.
(1N.ii) I will not marry, but I intend to marry.

Are these absurd to judge or to assert?

Again, the omissive case splits off from the others. (1O.ii) does not seem absurd to

[153] Certainly the second conjuncts do not necessarily imply that the dying will be unintentional; there's time to form an intention before the death, you might think. I return to this point later.

[154] In *Intention*, Anscombe provides verbs denoting actions which must be voluntary or intentional (1957, p.85). These are: telephoning, calling, groping, crouching, greeting, signing, signaling, paying, selling, buying, hiring, dismissing, sending for, marrying, and contracting. I don't agree with all of her choices, but I would like to retain the label "Anscombe verbs" to refer to just those verbs which denote actions that are constitutively intentional (not merely constitutively voluntary), and manifestly so. Compare Bennett (1988) on "intention-drenched verbs" (pp.205-6) and Michael S. Moore (2010) on "intentionally complex" verbs (p.174). I'm grateful to Juan Piñeros for highlighting these sources.

judge or to assert. The problem here is the future tense of the first conjunct. Constitutively, you need to have the intention to marry *when* you marry; but since we're talking about the future here, you still have time to form that intention to marry before you get to the big day. By those lights, it's perfectly consistent to marry in the future, but not now intend to marry. It also seems that you can recognize that much about yourself, all at once; and that is something that you could reasonably assert to someone else. For these reasons, neither judging nor asserting (1O.ii) seems absurd.[155]

As for (1C.ii) and (1N.ii): the same commitments that mattered for (1C.i) and (1N.ii) play the same role here as they did above. Insofar as it is possible to have, and to bring together, a belief that you will do something and an intention not to do it—or a belief that you will not do something and an intention to do it—you may be able to assert and judge (1C.ii) and (1N.ii) without absurdity, if not without oddness.

The use of an Anscombe verb, in short, doesn't seem to change much. Perhaps it adds to the oddness of asserting (1C.ii), since marrying in the future involves forming an intention you now assert you do not have. But, as with (1O.ii), you have time to form that intention before the big day: the first conjunct suggests the relevant action is in the future.

Once again, the unspecific futurity of the first conjunct is making trouble here. In all instances of (1O) – (1N) we have considered so far, there is a **temporal window** between the time one is said to have (or lack) some intention and the time one fails to act upon it (or acts upon an intention one now lacks). This temporal window allows for a number of important changes: decisions can be abandoned; obstacles can arise; and willpower can simply weaken. As long as this temporal window remains, it will be difficult to find sentences that are undeniably absurd to assert. Note, also, that there is no such temporal window in the case of Moorean absurdities for belief.

How do we need to amend our candidate Moorean absurdities for intention in order to do away with the temporal window? What we need to ensure is that there isn't a gap between the action in question and the self-attribution of the corresponding intention. That requires us to do away with the unspecific future tense of the first conjunct. Consider, then, the following:

(1O') I am Φ-ing now, but I don't intend to Φ.
(1C') I am Φ-ing now, but I intend not to Φ.
(1N') I am Φ-ing now, but I intend to Φ.[156]

Would this help? Consider some instances involving Anscombe verbs:

[155] There might be some oddness here, but I don't think it rises to the level of absurdity. You might question how the speaker knows that she will marry, without having the intention—but there are plenty ways of predicting your future, even accurately and reliably predicting your future, without going through your own intentions. You might assert (1O.ii) when you know full well, given your compulsion to repeat your (married) siblings' choices—that you *will* marry, but you have not yet formed the intention to marry. Perhaps you need to find the right person first to form a genuine intention: you don't intend to *marry*, simpliciter, but you know you will one day intend to marry *some particular person*, de re.

[156] You might alternatively close the temporal window by using one of the following conjuncts as the first: "I Φ" or "I will Φ now." All of the following points still apply, though.

(1O'.i) I am marrying now, but I don't intend to marry.
(1C'.i) I am marrying now, but I intend not to marry.
(1N'.i) I am not marrying now, but I intend to marry.[157]

By eliminating the temporal window and using Anscombe verbs, we have identified genuine absurdities in (1O'.i) and (1C'.i)—but not satisfiable ones. It's impossible to marry while you don't intend to, *or* while you intend *not* to. (1N'.i), on the other hand, is not absurd at all. These are not good candidates for Moorean absurdities for intention.

These problems arise because "marry" is an Anscombe verb—a verb that names a constitutively intentional action. Consider, then, instances of (1O') – (1N') *without* Anscombe verbs:

(1O'.ii) I am dying now, but I don't intend to die.
(1C'.ii) I am dying now, but I intend not to die.
(1N'.ii) I am not dying now, but I intend to die.

Curiously, putting the first conjunct into the progressive present tense seems to make the omissive and commissive versions *less* problematic than they were in the future tense: they now do not describe finished events, and do not entail that these events will be completed; and so someone who judged or asserted (1O'.ii) or (1C'.ii) could easily believe the first conjunct without seeing it as a closed, determinate matter that she will die. The problem of forced interpretation returns with a vengeance: it is easy to read the first conjuncts of (1O'.ii) and (1C'.ii) as describing an unintentional event in progress. As for (1N'.ii): this version, subject to a new temporal window, is not absurd at all.[158]

There is one last hope here. Some philosophers think "shall," unlike "will," expresses a contemporaneous intention rather than mere prediction.[159] If that's true, then the following should be Moorean-absurd:

(1O") I shall Φ, but I don't intend to Φ.
(1C") I shall Φ, but I intend not to Φ.
(1N") I shall not Φ, but I intend to Φ.

This would be an attractive solution to the problem of forced interpretation that doesn't rely on Anscombe verbs. However, "shall" is notoriously difficult to interpret.[160] Using

[157] To me, the unspecific "marry" seemed marked now that we have the time index in the first conjunct—whenever you intend to marry *now*, don't you intend to marry a particular person, *de re?*—so please excuse this additional adjustment.
[158] You might make a more absurd version in "I am not dying now, but I intend to die now." The interpretation of this sort of judgment, or this sort of assertion, seems to me to be subject to the same problems and controversies as the first examples in this section: (1O.i), (1C.i), and (1N.i). So do the versions that put the first conjunct into the immediate future, rather than into progressive present tense: "I will die now, but I don't intend to die," and so forth.
[159] See, e.g., Goldstein (1993).
[160] See the *OED* entry on "shall." One representative excerpt: "In the first person, *shall* has, from the early Middle English period, been the normal auxiliary for expressing mere futurity, without any adventitious notion … [either] (a) of events conceived as independent of the speaker's

"shall" will only offer a solution to the problem of forced interpretation if it is *impossible* to interpret "shall" properly here as anything *but* an expression of intention. But that is not impossible.[161]

Have we made any progress so far? We have not, perhaps, successfully identified actual Moorean absurdities for intention—at least not any that might serve as any kind of starting points for philosophical inquiry. The examples discussed above are, at best, subject to a number of delicate controversies.

However, we have come to understand better the challenges involved in formulating Moorean absurdities for intention. When the problem of forced interpretation arose, we tried to solve it with Anscombe verbs—and were left with the problem of the temporal window. In fixing that problem, we inadvertently produced unsatisfiable or commonplace sentences. These three problems—of forced interpretation, of the temporal window, and of unsatisfiability—are surprisingly recalcitrant.

Is there any way to avoid them? Let's try another strategy, one that involves a different kind of expression of intention.

3.2. Φ-ing is the thing to do

Gibbard (2003) argues that a practical fragment of ordinary language requires an expressivist analysis. To say that "Φ-ing is the thing to do," he writes, is just to express a plan—for our purposes, an *intention*—to Φ.

Gibbard's view suggests that the following are Moorean absurdites for intention:

(2O) Φ-ing is the thing to do, but I don't intend to Φ.
(2C) Φ-ing is the thing to do, but I intend not to Φ.
(2N) Not Φ-ing is the thing to do, but I intend to Φ.

With these schemas, the problem of the temporal window couldn't arise: in the right kind of utterance of an instance of (2O), for example, you just would *express* an intention at the same time as you *describe* yourself as lacking it.[162] Nor could forced interpretation

---

volition. (To use *will* in these cases is now a mark of Scottish, Irish, provincial, or extra-British idiom.) … [or] (b) Of voluntary action or its intended result. Here *I* (*we*) *shall* is always admissible exc. where the notion of a present (as distinguished from a previous) decision or consent is to be expressed (in which case *will* must be used). Further, *I shall* often expresses a determination insisted on in spite of opposition, and *I shall not* (colloq. *I shan't*) a peremptory refusal." Here's a recipe for an enjoyable night in: print the entry; pour yourself a cup of tea; peruse all 43 pages; and marvel at the complexity of the English language.

[161] A chilling example can be found in a letter from King Henry VIII to Anne Boleyn. The letter was written in 1527—nine years before he ordered her execution. He wrote: "I beg to know expressly your intention touching the love between us. Necessity compels me to obtain this answer, having been more than a year wounded by the dart of love, and not yet sure whether I *shall fail* or find a place in your affection" (quoted in Sifferlin 2012, emphasis added). Here "shall" is best interpreted *not* as an expression of intention. The King is not unsure whether he intends to fail; he is unsure about what in fact will befall him.

[162] Since Gibbard's view is an expressivist one, we might not want to speak of *assertion* of instances of (2O), (2C), and (2N). As expressions of intentions, the first conjuncts of the schemas above aren't technically truth-apt, on Gibbard's view. But Gibbard takes himself to have the

present a problem here.

However, "Φ-ing is the thing to do" may not be best understood as an expression of an intention to Φ. It does not seem absurd to my ear to assert any instance of (2O) – (2N). Any such assertion could describe a situation in which one's intentions fail to align with what one considers to be 'the thing to do.' Whether we understand that in terms of what is rationally required, what is morally obligatory, or in some other way, the fact is that your intentions can fail to align with 'the thing to do'—and you yourself can recognize and express that fact.

3.3. I know I will Φ

Perhaps we can exploit the link between knowledge and intention to construct new candidate Moorean absurdities:[163]

(3O) I know I will Φ now, but I don't intend to Φ.
(3C) I know I will Φ now, but I intend not to Φ.
(3N) I know I will not Φ now, but I intend to Φ.[164]

Consider instances of the above without Anscombe verbs:

(3O.i) I know I will die now, but I don't intend to die.
(3C.i) I know I will die now, but I intend not to die.
(3N.i) I know I will not die now, but I intend to die.

Here the problem of forced interpretation returns. It's not absurd to assert (3O.i), and you can still assert (3C.i) or (3N.i) to describe thwarted intentions.

Instances with Anscombe verbs seem more promising:

(3O.ii) I know I will marry Elena now, but I don't intend to marry Elena.
(3C.ii) I know I will marry Elena now, but I intend not to marry Elena.
(3N.ii) I know I will not marry Elena now, but I intend to marry Elena.

---

resources to make sense of a kind of speech act relevantly like assertion, and so he could still distinguish between those kinds of utterances of instances of (2O) – (2N) that are absurd, and those that are not.

[163] One potential explanation of the absurdity of asserting Moorean absurdities for belief has to do with the norms of assertion—and one view of the norms of assertion takes those norms to involve *knowledge* of what you assert. If this is your view, then considerations related to these sorts of explicitly knowledge-based candidates for Moorean absurdities for intention will be closely related to the considerations we saw in the previous section.

[164] I'm grateful to Susanna Siegel for drawing my attention to these ideas. I've time-indexed the first conjuncts of these to get the problem of the temporal window out of the way. Also note: other epistemic notions can be used to construct similar candidate sentence schemas. Consider, for example, "I'm certain I will Φ now, but I don't intend to Φ;" "I will definitely Φ now, but I intend not to Φ;" "I understand I will not Φ now, but I intend to Φ;" etc. Also see Yalcin (2007) for a discussion of the curious behavior of apparently Moorean-absurd sentences with epistemic operators (e.g. "it's raining but it might not be raining") in embedded contexts.

Once again, the cost of using Anscombe verbs is inconsistency in some cases, and banality in others. Since knowledge is factive, neither (3O.ii) nor (3C.ii) can be true; and (3N.ii) is unremarkable.[165]

3.4. I must Φ, I cannot Φ

You cannot intend to do what you believe you cannot do; and you cannot intend *not* to do what you believe you cannot avoid doing.[166] Consider, then, the following:

(4O) I must Φ, but I don't intend to Φ.
(4C) I must Φ, but I intend not to Φ.
(4N) I cannot Φ, but I intend to Φ.[167]

These schemas provide very good candidates for Moorean absurdities for intention, as long as "must" or "cannot" is interpreted with the force of *determination* rather than any moral, pragmatic, or other *obligation*. Let's walk through the reasons to think that these are good candidates for Moorean absurdities for intention with reference to the necessary conditions of Moorean absurdities for intention mentioned in the introduction.

First, instances of any of them can be true, for the same reason that instances of (O) – (N) can be true: people's beliefs can be mistaken or incomplete. Take (4N). Although you cannot intend to do what you *believe* you cannot do, you *can* intend to do what you *in fact* cannot—if you lack the relevant belief.[168]

Second, instances of (4C) or (4N) are absurd to assert even *without* time-indexing and Anscombe verbs:

(4C.i) I must die, but I intend not to die.
(4N.i) I cannot die, but I intend to die.

The absurdity of these instances implies that the problem of forced interpretation and the problem of the temporal window do not arise for (4C) and (4N). This nice feature is shared by the original Moorean absurdities for belief.

(4O) is a little fussier. Several of its instances are fine to assert:

---

[165] To reiterate a point made in a footnote above: inserting 'now' at the end of the second conjuncts as well doesn't help. Here, this move would render (3N.ii) inconsistent, and it wouldn't affect the inconsistency of (3O.ii) and (3C.ii).

[166] See, e.g., Davidson (1980), pp.100-1.

[167] Other notions that get at ideas of necessity or determinacy can be used to construct similar candidate sentence schemas. Consider, for example, "Φ-ing now is unavoidable, but I don't intend to Φ;" "I it is absolutely determined that I Φ now, but I intend not to Φ;" "It is impossible for me to Φ now, but I intend to Φ;" etc. It is important, however, that the *must* is one of factual inevitability, not moral or legal obligation (or any other flavor, for that matter).

[168] There are, of course, restrictions on these points concerning how *manifestly* something is impossible to do. Some of those restrictions are captured by the above restriction of the range of the variable Φ to exclude descriptions of actions like "something impossible to do." You can't, I take it, really *lack* the belief that something that's impossible to do is impossible to do.

(4O.i) I must die, but I don't intend to die.
(4O.ii) I must die now, but I don't intend to die.
(4O.iii) I must marry Elena, but I don't intend to marry Elena.

Each such instance has *no* time-indexing, no Anscombe verbs, or neither—so each is vulnerable the problem of the temporal window, the problem of forced interpretation, or both. When asserted, (4O.i) and (4O.ii) both seem like descriptions of unintended inevitabilities, since their second conjuncts force this interpretation; an assertion of (4O.iii) can express that one hasn't *yet* formed an intention that one *must* have later on.

Yet consider the following instance, time-indexed with an Anscombe verb:

(4O.iv) I must marry Elena now, but I don't intend to marry Elena.

This sentence can be true when the subject lacks the relevant belief, but it is absurd to assert it. In assertion, it could not describe an unintended inevitability, since marriage is constitutively intentional—and it could not express that one hasn't yet formed an intention one must have later, since the action must happen *now*. This is a good candidate for a genuine Moorean absurdity, just like (4C.i) and (4N.i). Note that (4C) and (4N), as well as instances of (4O) with both kinds of adjustments, are absurd to assert *only* in first-personal present-tense.

Perhaps it should not be surprising that (4O) is fussier than (4C) and (4N). There is a way to make sense of this discrepancy between the forms—a discrepancy that doesn't exist between forms of Moorean absurdities for belief.

Beliefs and intentions are rationally constrained in distinct ways. Having intentions that are manifestly jointly unsatisfiable—like the intention to Φ and the intention not to Φ—is rationally unsustainable, but there are no requirements to form intentions to respond to the way the world is, as there are for beliefs. Roughly speaking, beliefs aim to capture the way the world is, whereas intentions aim to change it. The fact that the world is—or will be—a certain way does not demand that you form an *intention* to bring that about; even knowing that makes no such constraint. Only when you believe explicitly that you must *now* do something *intentionally* are you constrained to form that intention.[169] I will return to this point later in this chapter, in arguing that none of the candidates discussed here are *genuine* Moorean absurdities for intention.

3.5. Speech acts

Acting intentionally may be a more fundamental an expression of that intention than saying that you *will* so act. Can we use this fact to construct yet more Moorean absurdities for intention?

Consider performative speech acts that can be conjoined with claims about intentions. For performative verbs[170] Φ, saying *I Φ* is a way of Φ-ing. For such verbs, the following

[169] Some would likely argue that knowing that much is tantamount to having the intention.

[170] I leave this term undefined in this paper, since its proper definition is a matter of philosophical controversy. I will say that performative verbs must meet at least the following conditions: for any performative verb Φ, saying *I Φ* is a way of Φ-ing. Performative verbs of this kind include "promise," "order," "apologize," and many more. See, e.g., Austin (1962) and Searle (1989).

schemas may provide candidate Moorean absurdities for intention:

(5O) I Φ, but I don't intend to Φ.
(5C) I Φ, but I intend not to Φ.
(5N) I do not Φ, but I intend to Φ. f

However, not all performative verbs Φ will work here: performative *Anscombe* verbs, for example, will not. Consider promising. You constitutively cannot promise without intending to promise. For that reason, the following are not Moorean absurd:

(5O.i) I promise, but I don't intend to promise.
(5C.i) I promise, but I intend not to promise.
(5N.i) I do not promise, but I intend to promise.

(5O.i) and (5C.i) are each inconsistent. (5N), on the other hand, is unremarkable.

Some performative *non*-Anscombe verbs, however, offer more interesting instances:

(5O.ii) I defy you, but I don't intend to defy you.
(5C.ii) I defy you, but I intend not to defy you.[171]

These are good candidates for Moorean absurdities for intention.

First: both (5O.ii) and (5C.ii) can be true. You can defy someone without intending to, or while intending *not* to defy her; consider misinterpretations of commands.

Second: each is absurd to assert—insofar as asserting either one is, by way of the first conjunct, *a way of* defying the addressee.[172] This would not be the case with non-performative verbs: what is special and relevant about the performative verb "defy" here is precisely that *saying* "I defy you" is a way of doing just that.[173]

Third: the absurdity in such assertion vanishes in non-first-personal, or non-present-tense, versions of instances of (5O) and (5C), even with performative non-Anscombe verbs. It's not at all absurd to assert, say, "she defies you, but she doesn't intend to defy you," or "I defied you, but I intended not to defy you."

Note that no problem of forced interpretation—nor any problem of the temporal window—arises for these cases. This is an important parallel between these cases and our

[171] Other performative verbs that aren't Anscombe verbs might include: "praise," "defer," "honor," "volunteer," "insult," and "allow." These will probably be variously controversial.

[172] Assertoric accounts of performative speech acts are controversial: see, e.g., Searle (1989). For our purposes, this complication will not much matter: it is the contrast with hypothetical utterances of claims like (5O.ii) and (5C.ii) that matters much more than whether we give an assertoric account of performative speech acts.

[173] The following is not true: whenever saying *I Φ* is a way of Φ-ing, Φ is a verb of the kind that would make an instance of (5O) or (5C) absurd. Consider the following counterexample: someone might get upset whenever the word 'upset' is used. Not knowing this, I might say to her "I upset you, but I don't intend to." Now, saying "I upset you" in this instance *is* a way of upsetting my addressee—but it's not absurd for me to make that assertion. That's because "upset" is not a performative verb; and so we see that there must be further necessary conditions on performative verbs that I do not discuss here, for the purposes of space.

original Moorean absurdities for belief.

Are there any such Moorean absurdities in *negative-commissive* form? Importantly, the values for Φ that make instances of (5O) and (5C) absurd will not always work to make instances of (5N) absurd as well. Consider:

(5N.ii) I do not defy you, but I intend to defy you.

Even when saying *I Φ* is a way of Φ-ing, it's not guaranteed that saying *I don't Φ* is a way of *not* Φ-ing. Saying "I do not defy you" is *not* a way of *avoiding* defying the addressee (although it certainly is *not* a way of defying the addressee).

Any verb Φ that will make a Moorean-absurd instance of (5N) must meet at least these necessary conditions: saying *I do not Φ* is a way of not-Φ-ing; and not-Φ-ing is not *itself* constitutively intentional. One such verb is *accept*:

(5N.iii) I do not accept it, but I intend to accept it.

(5N.iii) seems to have satisfiable truth-conditions, and it seems absurd to assert—but only in its first-personal, present-tense form. (5N.iii), then, is a good candidate for a Moorean absurdity for intention.

Along similar—but not quite the same—lines, it is worthwhile to pause and consider a schema discussed by Searle and Vanderveken (1985) as an example of a Moorean-absurd sentence:

(5O') I promise to Φ, but I don't intend to Φ.

Promising is special among speech acts as a sort of public commitment to doing something in the future. As such, the performative verb "promise" can be used to provide a somewhat direct expression of intention in (5O'). It is plausible that similar verbs that involve almost identical performative speech acts—including "pledge" and "commit"—can be used in just the same way. Alongside (5O'), then, it is instructive to consider commissive and negative-commissive versions of the same sort of sentence:

(5C') I promise to Φ, but I intend not to Φ.
(5N') I do not promise to Φ, but I intend to Φ.

Are instances of (5O') – (5N') Moorean absurdities for intention?

The first point to consider is whether instances of these schemas can be true. It seems straightforward that they can be. Although *promising* is plausibly constitutively intentional—you cannot promise without intending to promise—it's not constitutive of promising to Φ that you intend to Φ. Promises can be insincere in that way. That implies that instances of (5O') and (5C') can be true.

Instances of (5N') can quite clearly be true, as it is certainly no condition on *not* promising to Φ that you intend to Φ. For that very reason, though, it should be clear that no instance of (5N') will be absurd to assert: you might reasonably assert something like

(5N'.i) I do not promise to marry, but I intend to marry

in a case in which you have enough confidence that you *can* marry to reasonably intend to marry, but not enough confidence to promise to your addressee that you will succeed.

How about instances of (5O') and (5C')? In cases with Anscombe verbs, instances of either seem robustly absurd to assert:

(5O'.i) I promise to marry, but I don't intend to marry.
(5C'.i) I promise to marry, but I intend not to marry.

Notice that, as with instances of (5O) – (5N) above, no time-indexing is necessary here: the speech act directly implies the presence of an intention at the time of utterance, and thus automatically avoids any problem of the temporal window. It's also important that non-first-personal or non-present-tense versions of these instances are not absurd to assert. These features imply that we have come upon another reasonable candidate for Moorean absurdities for intention—at least when Anscombe verbs are filled in for Φ.

However, something more interesting happens in cases with non-Anscombe verbs:

(5O'.ii) I promise to die, but I don't intend to die.
(5C'.ii) I promise to die, but I intend not to die.

These two sentences seem to differ in their absurdity. Sufficient confidence that something *will* happen in the future—like your dying—may well lead you to assert (5O'.ii) without absurdity: forced interpretation leads us to understand the promise along the lines of an assurance of certainty rather than along the lines of a decisive commitment on the part of the speaker to bring something about. (5O'.ii) and other instances of (5O') *without* Anscombe verbs, then, are not good candidates for Moorean absurdities for intention.

However: no such reading is available for the commissive variant. To read the first conjunct of (5C'.ii) as an assurance of certainty would be to read it like the first conjuncts of instances of (4C)—all of which instances we took to be absurd. Since no such interpretation is available in the case of (5C'.ii) and other instances of (5C') without Anscombe verbs, we can safely conclude that the use of Anscombe verbs is immaterial to the absurdity of instances of (5C').

It seems, then, that no instances of (5N') are absurd; that instances of (5O') *with* Anscombe verbs are absurd; and plausibly *all* instances of (5C') are absurd.

The kinds of absurdities discussed in this subsection use speech acts as expressions of intention. That implies that these kinds of contents are absurd to *assert*, but may not be absurd to judge at all. This is an important disanalogy with the Moorean absurdities for belief discussed in Chapter 3.

3.6. Classifying absurdity

In light of the discussion above, it seems that instances of all of the following are good *candidate* Moorean absurdities for intention:

(4O') I must Φ now, but I don't intend to Φ

- *where Φ is an Anscombe verb*

(4C) I must Φ, but I intend not to Φ.

(4N) I cannot Φ, but I intend to Φ.

(5O) I Φ, but I don't intend to Φ

- *where Φ is a performative non-Anscombe verb*

(5C) I Φ, but I intend not to Φ

- *where Φ is a performative non-Anscombe verb*

(5N) I do not Φ, but I intend to Φ

- *for some verbs Φ such that: saying* I do not Φ *is a way of not Φ-ing; and not Φ-ing is not itself constitutively intentional*

(5O') I promise to Φ, but I don't intend to Φ

- *where Φ is an Anscombe verb*

(5C') I promise to Φ, but I intend not to Φ.

However, to say that these are good *candidates* for Moorean absurdities for intention is not yet to say that these *are* Moorean absurdities for intention. I have already noted above some important disanalogies between the (5) forms and Moorean absurdities for belief: *judgments* of contents of any of the (5) forms will not be absurd. These forms are constructed out of performative speech acts, and so they are only absurd to assert, not to judge silently to oneself. Insofar as it is *essential* to Moorean absurdities that they be absurd to judge, these will not be Moorean absurdities for intention.

What does it take for some content to be a *Moorean* absurdity, rather than some other kind of absurdity? That is a difficult question to answer. Its answer might be indeterminate. I identified Moorean absurdities for belief by describing their forms instead of giving necessary and sufficient conditions on some content's being a Moorean absurdity for some mental state or mental act *M*. A simple transfer was required to find Moorean absurdities for judgment that have almost precisely the same forms.

But asking the same question about Moorean absurdities for intention can be done on several levels. It is not clear what could settle which level of generality we need to use in order to identify whether the absurdities discussed above are genuinely *Moorean* absurdities. The thing to do, then, is just to clarify what is meant by "Moorean absurdity" in this context, and then to determine whether the absurdities concerning intention meet the conditions on the notion we are using in this context.

Here, I take a Moorean absurdity for intention to be a content that is absurd to judge *or* to assert, and one whose absurdity is explained in precisely the same way as the absurdity of judging or asserting something of the form of (O), (C), or (N).

In the previous chapter, I explained why it was absurd to judge or to assert any Moorean content involving a belief attribution. That explanation appealed to a central case in which it is *impossible* to judge any content with a Moorean form. It is impossible to judge any content with a Moorean form when you are engaged in judgment intentionally. That is because engaging in judgment intentionally, in a context in which you are considering what you believe, gives you all you need to make it the case that you recognize what you believe just in judging the first conjunct of the absurdity. A judgment that *p* (or a judgment that it's not the case that *p*) can also *be* a self-attribution of a belief

with the same content, *p* (or the content that it's not the case that *p*). It matters here that judging that *p* is *sufficient* for having a belief that *p*, at least at the moment of judgment. And so it matters that a judgment of any of those Moorean absurdities with conjunctive contents involves that judgment itself.

What would be needed for there to be Moorean absurdities for intention whose absurdity (in judgment or assertion) is explained in just the same way? There would have to be some way of concatenating, in one thought, your *decision* with a judgment about your intentions on the same matter. There would then have to be some central case involving an intentional context in which it is absolutely impossible to think the concatenated thought that combines together, in one, the decision and the self-attribution.

The overarching challenge here was to find some content one could *judge* that would amount to making a decision to Φ. We had to find some relevant judgment because the conjuncts of a judgment must themselves be contents you judge. Above, though, we could not find a direct expression of a decision to Φ. Judging that you will Φ is not the same as deciding to Φ. Judging that Φ-ing is the thing to do is also not the same as deciding to Φ. We could not use these kinds of expressions to capture decision. If these were the only options available, we would have no Moorean absurdities for intention.

However, there are judgments that you can make that bear on whether or not *to* decide to Φ. Whether or not you *can* Φ matters to your decision whether to Φ. This last fact informed the production of forms (4O'), (4C), and (4N) as listed above. Moreover, there are ways of expressing your decision in different ways—e.g. by performing an intentional speech act. That fact contributed to the production of forms (5O), (5C), (5N), (5O'), and (5C') as listed above. Can we use these expressions to motivate an argument that it is *impossible* to judge something of these forms?

We cannot. This is easy to see in the case of (5)-forms, because these are not absurd to judge at all—whether or not you are judging intentionally.

Consider, then, the case in which you judge something with a (4)-form intentionally. If we are to demonstrate that these are forms of genuine Moorean absurdities—as I have defined that class—it must be impossible to judge any content of any of these forms when you are judging intentionally. For that to be impossible for the same reasons as in the belief cases, it would have to be that having practical knowledge of your *judging* as such was sufficient to enrich the judgment of the first conjunct into a self-attribution of an intention. However, your *judging* intentionally does not give you practical knowledge of the right kind of mental action to embed into a transparent self-attribution of an intention. You would have to know what you are doing as *deciding what to do* in order to self-attribute an intention in making a decision. That is how we would get the conflict up and running with the second conjunct of these judgments, and that itself is required to explain the sheer impossibility of judging something of the relevant form.

There is not a core case of impossible judgment there to anchor a proper explanation of the absurdity of judging or asserting something of one of these forms. Our failure to find an inescapable manifestation or expression of decision in the form of judgment ensures that none of these candidate Moorean absurdities for intention qualify as genuine Moorean absurdities for intention—cases in which the *same* absurdity arises as for belief.[174]

[174] This is not to say that the absurdities listed above are not philosophically revealing, and it is

Our failure to find genuine Moorean absurdities for intention suggests that there is no general analysis of what it is to decide to Φ that reduces decision to some judgment that something is the case. The search for Moorean absurdities for intention turned into a search for an expression or manifestation of decision that could be concatenated with a judgment to form a conjunctive judgment of some plausible format. Since any conjunct of a conjunctive judgment is something you judge, that just was a search for a judgment that manifests or expresses decision.

Decision is the occurrent mental act that shares an attitudinal aspect with intention, and judgment is the occurrent mental act that shares an attitudinal aspect with belief. For this reason, the results of this chapter also suggest that intention cannot be analyzed in terms of belief, despite philosophical attempts to produce some such analysis.[175] To have a belief is to take a fundamentally distinct stance towards the world than to have some intention. This can be captured in the rough thought that beliefs try to capture the world as it is, while intentions try to shape the world into something it is not yet.

Conclusion

In this chapter I demonstrated that the mental state of intention, and the mental act of deciding what to do, each meets the conditions on transparent self-knowledge. I explained how you can transparently self-attribute an intention or a decision, and I explained how those transparent self-attributions constitute authoritative knowledge of what you intend to do or what you are decided on doing. In order to explain the first-person authority that attaches to self-attributions of intention and decision more generally, we need to note an important fact that has arisen in previous chapters as well. We assume of one another that we use the transparency method to self-attribute intentions and decisions by default, and we only use other methods when the transparency method is not available. This will become important again in Chapter 6.

Intention and decision also meet additional conditions that imply that there should be Moorean absurdities involving intention attributions and decision attributions. There are indeed conjunctive contents involving intention attributions that are absurd to assert or to judge—and only in the first-person, present-tense case. These contents share some important features with Moorean absurdities for belief. But the absurdity in judging or asserting these contents does not arise in quite the same way as the absurdity—and sometimes, impossibility—of asserting or judging the corresponding contents about belief. They are not, then, Moorean absurdities for intention in that sense. And the fact that they cannot be Moorean absurdities for intention in that sense suggests that intention is not any form of belief—and decision to act is not any form of judgment.

---

not to rule out that there might be some broader class of absurdity that captures both Moorean absurdities for belief and the forms for intention we have recognized. It is just to clarify that the absurdities we have found do not meet the substantive conditions I am interested in for the purposes of this chapter. For more on broader classes of absurdity, see Wall (2012) and Williams (2014), and especially Sorensen (1988) and Green and Williams (2007).

[175] See, e.g., Velleman (1989).

Chapter 5. Groundlessness and the First-Person Perspective

In this chapter I would like to bring out a similarity between the way in which we self-ascribe beliefs and the way in which we attribute beliefs to others. We make epistemically groundless ascriptions of beliefs not only to ourselves but also to others. To gain any intentional knowledge of others' actions, behavior, and mental states by observation, we must ascribe beliefs to other people groundlessly. By "intentional knowledge" here, I mean knowledge of those intentional attitudes and actions in intensional terms—the terms that accurately describe how the subject sees herself and the world with which she interacts.

Why care that we make groundless ascriptions of belief not only to ourselves, but to others? There is at least one simple motivation: the literature on self-knowledge often cites as a fact the claim that *groundless* ascription of belief is special to the first-personal perspective, often along the way to an explanation of first-person authority. However, we don't actually need to accept this claim to make sense of first-person authority.

More generally, thinking of groundlessness as special to first-personal ascriptions of belief is to miss the important fact that some groundless judgment is crucial to intentional knowledge. I think that this point is worth considering in its own right.

What does "groundless" mean, in this context? For a judgment or a belief to be groundless is to say that its subject (the one doing the judging or believing) has no reason for that judgment or belief that justifies it. That is not necessarily to say that the judgment or belief is entirely unwarranted. You could have entitlement for some groundless judgment or belief of yours. Some people think that "I exist," in the context of the cogito, is like that: groundless, but warranted by way of entitlement.[176] Additionally, to say that a judgment or belief is groundless is not to say that there are no causal reasons that explain your making that judgment or your forming (or sustaining) that belief. Consider a silly example: a whack to the head could (conceivably, at least) cause you to believe that the Pope is on the moon right now. You could have no reason for this belief in the sense that there is nothing accessible to you that justifies that belief, epistemically speaking. But there is certainly a reason you have that belief in the causal sense: the reason is that you were whacked on the head in a particular way. This is a case of a belief that is, overall, unwarranted: you have neither justification nor entitlement for the belief.

There is something internalist in this understanding of epistemic grounds: to say a judgment or belief of yours is groundless is to deny that *you have* any reason for it. To argue for a judgment's or belief's groundlessness, then, the candidate grounds we must rule out are things that are, broadly speaking, *available* to the subject.

I speak of *third-personal ascriptions* and *other-ascriptions* of belief in this paper interchangeably, though it is worth noting upfront that you can make third-personal ascriptions of beliefs (and all sorts of other states) to yourself too: imagine seeing yourself from an angled overhead view on a CCTV camera, not knowing it is *you* ("me").

[176] See, e.g., Burge (1996).

1. Intentional interpretation, observation, and underdetermination

The first step on the way to recognizing that some third-personal ascriptions of belief are *groundless* is to understand that giving any *observed behavior* any intentional description also involves making a number of further *implicit* belief ascriptions to the individual to be so described.[177]

Behavior itself, described in purely *extensional* terms, underdetermines its correct intentional description.[178] This fact has been discussed, most famously, by Davidson, Quine, and Kripke.[179] The point runs through Davidson's work on radical interpretation, Quine's work on the indeterminacy of translation, and Kripke's work on what he calls the "Wittgensteinian paradox," and it takes distinct forms in each such discussion, but the key point in question can be summarized thus: any purely extensional description of observable behavior *underdetermines* its intentional nature—what beliefs or desires that behavior expresses, what intentions it fulfills, or what it *means*. Thus, to figure out what someone *meant* by 'plus' in the past, it is not enough to observe what she *actually* did (e.g. write "125" on the board after "68+57=") (Kripke). To determine what a native of a completely unfamiliar culture *means* by "Gavagai," it is not enough to observe him pointing at rabbits and uttering this word, over and over again; for those observations are consistent with his meaning any number of things, including "rabbit," but also "undetached rabbit parts" and "temporal rabbit-slice" (Quine). As Davidson puts it, "behavioural or dispositional facts that can be described in ways that do not assume interpretations, but on which a theory of interpretation can be based, will necessarily be a vector of meaning and belief."[180] There are infinite ways of attributing belief and meaning to an individual that make sense of her behavior. In order to settle on what a person *means*, Davidson argued, we need to hold fixed in some way what she *believes*.

Davidson, Quine, and Kripke raise this point in contexts that we might call epistemically extraordinary—contexts involving *radical* interpretation, *radical* translation, or *skepticism* about meaning. In such contexts what is available as evidence is particularly poor, or what counts as knowledge is particularly circumscribed. What is

---

[177] An intentional description of behavior is a description of that behavior in terms of intentional attitudes in intensional terms. This description could just be a description of behavior as an *intentional action*, but it could also involve saying a lot more about the observed subject's desires, beliefs, and so forth.

[178] I do not here mean, by "observable behavior," what Lewis (1974), in writing about radical interpretation, means by "**P**"—that is, "the whole truth about Karl [for example] as a physical system ... [including] how Karl moves, what forces he exerts on his surroundings, what light or sound or chemical substances he absorbs or emits" (p.331). I am tempted to think that the arguments of this paper would still be sound if I *did* mean that by "observable behavior" (pace Lewis, p.334), but I want to avoid commitment on this point for now, and say rather that "observable behavior" just involves all those bodily movements that we ordinarily see, hear, feel, and so forth. I do here mean to borrow a different disambiguation of Lewis's: "There is an ambiguity in the term 'behavior'. Note that I am using it to refer to raw behavior – body movements and the like ... not to refer to behavior specified partly in terms of the agent's intentions" (p.338). This is what is meant by "behavior described in purely *extensional* terms."

[179] Davidson (1973/2001, 1974a/2001, 1974b/2001, 1991/2001), Quine (1960), and Kripke (1982).

[180] Davidson (1974a/2001), p.148.

important to this paper, though, is that a version of this underdetermination problem arises in epistemically *ordinary* contexts as well. Davidson in particular does not seem to have recognized the extent to which this problem extends to situations in which you understand the language of the person to be interpreted.[181] I want to illustrate that it does.

I'd like to demonstrate how the problem of underdetermination of intentional description applies in everyday contexts partly by considering a particular example. Imagine you have just sat down in a gym to take the SAT. A sixteen-year-old girl you have never seen before sits just to your right. About an hour into the test, you notice that she is tapping the tip of her pencil rhythmically on her desk. It's distracting and irritating. Why is she doing that? What is she doing that for? What does she mean by this? Or, simply: *what is she doing*? What you're looking for is an intentional description of her behavior, one that makes sense of what she takes herself to be doing in intensional terms.

It occurs to you that she might not even know she is doing it—that is, that she's not doing it *on purpose* at all, but perhaps as a matter of nervous habit. But then again, maybe she *is* doing it intentionally, even if it's out of anxiety: perhaps she's tapping the beat to her favorite power ballad, to get her pumped for this next math section. Maybe that's not it, though. Maybe she's in cahoots with another student around here, and she's tapping out a code that holds the answers to the multiple-choice questions. Maybe she's doing it to break the lead in her pencil, so she'll need to go up and ask for a new one, and then she can peek at other people's answer sheets on the way up to the proctors' desk at the front of the gym. Or maybe—and this is most irritating of all—maybe she's doing it *to annoy you specifically*. As things are now, you can't definitively rule out any of these options.

This is not a situation of radical interpretation or radical translation—indeed, this is not necessarily a situation in which you aim to interpret someone's *language* (except perhaps if she really is tapping out a code to her co-conspirator). Nor is it any kind of skeptical situation, where some intentional description of her behavior is under threat precisely for its intentional nature. This is just intentional interpretation in the wild.[182]

---

[181] It is not easy to pin down his position on this point, but see Davidson (1973/2001, 1974a/2001). Davidson writes: "There is a principled, and not merely a practical, obstacle to verifying the existence of detailed, general, and abstract beliefs and intentions, while being unable to tell what a speaker's words mean ... The absurdity lies not in the fact that it would be very hard to find out these things without language, but in the fact that we have no good idea how to set about authenticating the existence of such attitudes when communication is not possible" (1974a/2001, pp.143-4). The soft implication of comments like this is that communication *is* a way of authenticating the existence of such attitudes. And I do not mean to imply that it cannot be. But there is still always the possibility of deception, or of misunderstanding, and trusting a person's words—taking her speech at face value—involves making groundless implicit belief ascriptions that rule out the sorts of beliefs, desires, and intentions associated with such deception and misunderstanding. It may be pointed out that Davidson (1984/2001, 1987/2001) does, in explaining first-person authority, point out first- and third-personal asymmetries that have to do with failures in interpretation of speech behavior. But once again the possibility of failure here that Davidson recognizes is just the possibility of failing to speak *exactly* the same language as another person. I think, therefore, there is good reason to think that Davidson did not extend his concerns about intentional interpretation and the challenges involved to consider everyday situations that involve people who share exactly the same language (per impossibile?).

[182] Take "intentional interpretation" to be the activity of applying an intentional description to observed behavior.

You want to know what she's doing, and you don't have enough information to figure it out.

This is the crucial point: given just your observation of this student, tapping away at her desk, you don't have all the information you need to settle which of your hypotheses is correct. Your observation of her behavior *underdetermines* any intentional description of that behavior.

What more would you need to figure it out? A natural thought is that you need to know *what she's thinking*. This is an intuitive way of putting the point that you need to know more about her overall state of mind. And, as part of that, you need to know a little bit more about her *beliefs*. If you knew that she believed that tapping her pencil that hard would definitely break the lead, and that she *wanted* to break the lead, then you could easily ascribe to her the intentional action of *trying to break the lead in her pencil*. If you knew that she believed her friend Garth was around somewhere listening for Morse code signals indicating correct answers, and that she *wanted* to give Garth the answers in Morse code, you would know she was *giving Garth the answers*. And so on. The point is that part of what's missing is an understanding of this student's intentional set—her beliefs, her desires, and her intentions. If you could fix a few more intentional variables, you could better understand what she is doing. In particular, for our purposes, I'd like to highlight that you need to know a bit more about her *doxastic set*: her total belief state. Of course, knowing *that* while knowing nothing about her desires may not get you very far, but the belief part is just as crucial. Not only are "behavioural and dispositional facts ... a vector of *meaning and belief*," as Davidson put it: they are complex products of an entire intentional set, including a complex doxastic set.

This example is somewhat special in one way: you know close to nothing about the student, and so you lack the information you need to give her behavior an intentional description. Perhaps you suspend judgment (if not annoyance) with regard to what she's doing. But there is another way in which the example is a good representative example: it highlights the *general* dependence of intentional interpretation on further belief ascription (as well as further desire ascription, further intention ascription, and so on, but let's leave that aside for now). When you are observing another person you know very well—say, your best friend—you can interpret her in intentional terms much more easily than you can interpret the unknown student and her irritating tapping. But any given interpretation of her behavior in intentional terms will *still* depend on further knowledge you have about what she believes, what she wants, and so forth. The fact that those further belief ascriptions in cases of interpreting your best friend are more forthcoming, and more certain, than in the case of the unknown student, does not mean they are any less crucial to your interpretive activity.

Call such belief ascriptions—those which are necessarily involved in any intentional description of another person's behavior—*implicit* other-ascriptions of belief. I will argue that some such implicit other-ascriptions of belief are made *groundlessly*.

2. Groundlessness

Some implicit other-ascriptions of belief are made groundlessly. The most important point in the argument for this claim is the fact that not all such implicit other-ascriptions of belief can themselves be based on observation, on pain of regress. An argument by

reductio can demonstrate that much. Given that observation of behavior alone underdetermines intentional description, then to interpret any given person's behavior that you observe, you need to settle some further intentional facts about that person—including what she believes. Say that some behavior X is interpreted by way of intentional description I_X. But to give I_X as a description of X involves ascribing some other beliefs $B(I_X)$ to the agent in question as well. Where could those belief ascriptions come from? Well, let's say they came from observation of some other behavior, Y. Y itself underdetermines an intentional description, so the beliefs $B(I_X)$ could be ascribed based on observation of Y only given some intentional description I_Y of Y. As with all other intentional descriptions, though, I_Y will involve ascribing some other beliefs $B(I_Y)$ to the agent. Where do *those* ascriptions come from? If the answer were "from observation, and so on and so on, ad infinitum," then the particular intentional interpretation I_X at issue would never find firm footing—it could not be justified at all.[183] But I take it that often you *are* justified in giving another person's behavior intentional description of some particular form (I_X). So not all belief ascriptions can be made by observation.[184]

That is not *yet* to say that there are some implicit belief ascriptions that are groundless: perhaps there is some other way of understanding how you can make belief ascriptions such that those belief ascriptions are *grounded*, but not by way of observation. In particular, it may seem fruitful to recognize that you never actually face the sort of regress just rehearsed when you are trying to describe another person intentionally, and to think about why. Without the basis of observation to go on, how do you possibly ascribe these background beliefs? Precisely by assuming, of the person to be interpreted, that she has *true beliefs*. To make these sorts of assumptions is to follow what has sometimes been called "the principle of charity."[185] The reason that this has been understood—by Davidson in particular—as *crucial* to the activity of interpretation is precisely because, without this principle, you'd have nowhere to start: the regress would loom, and an infinitary set of intentional descriptions would open up before you, and you'd have no principled way of choosing between any two such descriptions.

I think it is fairly straightforward that we do in fact choose which implicit belief ascriptions to make to other people just by picking the *true* ones as much as is possible, while still making sense of others' behavior. I will take this for granted here. But the truth of such implicitly ascribed beliefs alone does not give those implicit belief ascriptions any particular *grounds*. p itself is not, in a perfectly general way, any good reason to think *S believes that p*. No valid deductive inference leads from p to any *S believes that p*. *S believes that p* is not the best explanation of p (far from it).

Could these implicit belief ascriptions be justified by induction over observed instances—instances that demonstrate that *people generally do believe true things*? If it is indeed true that all such observation requires such implicit belief ascription, then aggregated observations of further instances cannot themselves be justification for those very implicit belief ascriptions in future instances. To accept some such inductive justification would be to allow the practice to justify itself in a viciously circular way. For that reason, pointing to induced generalizations over observations of behavior cannot

[183] I suppose here I commit to rejecting infinitism about justification. I think that's all right.
[184] This regress argument also bears against Ryle's (1949) theory of self-knowledge, which takes it that *all* self-knowledge is observational. Cassam's (2014) picture may face the same challenge.
[185] See Lewis (1974).

explain the justification you have for implicit belief generalization in intentional interpretation of observed behavior.

Perhaps it's independently implausible, however, that we make such bold generalizations over such a broad class to serve our intentional interpretations of others. It is more plausible to think that we make more sensitive generalizations over restricted, but more meaningful, classes of people. There are finer-grained distinctions between people that affect whether or not they have beliefs that *you* take to be true: whether they have access to the same evidence as you, whether they are looking in the same direction as you are, whether they are currently conscious, whether they grew up in a socially progressive environment, whether they share some of your genes, and so forth. Perhaps inductive generalizations over these circumscribed groups of people are the source and grounds of the implicit ascriptions of true beliefs to others that we make in interpreting their behavior in intensional terms. The suggestion is that, rather than justifying all instances of the practice of *ascribing true beliefs to others implicitly* by induction to the crude generalization that *people generally believe true things*, each such implicit belief ascription is justified by way of a generalization like *social conservatives generally agree with me about abortion rights* or *my sister always has the right opinions about books* or even *people looking at me know what I'm wearing.*

There are several different points entangled in the suggestion as I've presented it, so let's try to tease these issues apart. One point to make is that sometimes these sorts of generalizations *do* ground—in a straightforward, inductive way—some particular implicit belief ascriptions to other people. Another thing to notice is that this fact *alone* does not yet disprove the thesis we set out to prove, namely that *some* implicit belief ascriptions to others are completely groundless.

Another important fact is that not all the generalizations you might use to ground particular implicit belief ascriptions to others are themselves inductively supported. To call someone a social conservative is to imply certain things about her attitudes towards, say, abortion, or marriage rights. A comparable generalization might take legal experts to have true beliefs about the law. Some descriptions like that just do have defeasible but non-inductively supported—perhaps *a priori* or analytic—implications about the doxastic set of the people to whom they apply. I'll return to this point below.

The most important point here, though, is that inductive support for any one of these generalizations that really does need inductive support to ground any belief ascription—e.g. the generalization that *people looking at me know what I'm wearing*—cannot get any such support without making some implicit belief ascriptions. Crucially, for these generalizations to help you form implicit belief ascriptions in a grounded way, they must be generalizations that relate some group of people to beliefs *intentionally ascribed*—or else the same problem of underdetermination that these generalizations are meant to help *avoid* would arise all over again. But in order to confirm any such generalization by way of observation, you would need to give the behavior of others that you observe an intentional description. That requires the same sort of interpretation under scrutiny here, which itself requires implicit belief ascription.[186]

[186] Here I am already setting aside further problems for these inductive generalizations. One further problem is this: if you really needed to have one of these generalizations in hand to ground any given implicit belief ascription, you'd need a *whole lot* of them (perhaps infinitely many), and your past observations would never be rich or multitudinous enough to provide much

Inductive generalizations of even a more sensitive and particular sort, then, will not be able to absolve us from the requirement of making *groundless* implicit belief ascriptions in intensional interpretation of other people. Even if there is some inductive practice that grounds *some* particular belief ascriptions—and there probably is—not all such belief ascriptions can be justified by induction "all the way down," for the same reason that they can't be justified by particular past ascriptions to individuals all the way down. The same regress from above looms all over again.

But how about those other generalizations—generalizations about legal experts, social conservatives, and so on? These non-inductively supported claims about groups of people (so grouped due to their doxastic properties) will certainly be useful in intentional interpretation of others whom you can knowledgeably (or at least, justifiably) describe as belonging to some such group. But sometimes—indeed, often—when you interpret others intentionally, you will have no such description in hand. Consider, again, the case of your irritating neighbor in the SAT. It is an unfortunate fact about American education that a description of someone as *sitting the SAT* implies nothing in particular about her doxastic or epistemic set. You cannot rely on such descriptions to ground *all* the many and various implicit belief ascriptions you have to make in intentional interpretation.

Is there any other way to understand yourself as having some *grounds* to ascribe beliefs to others implicitly in intentionally interpreting them? Perhaps you could have some broadly philosophical reasons, like transcendental reasons, that justify the implicit belief ascriptions that are not otherwise grounded. For example: you might think that a person must believe the external world exists in order to interact with it at all, and think that observation of behavior determines interaction with the world. If that's true, you have a priori transcendental reason to make the relevant implicit belief ascription to others of belief in the external world. It may be true that you have some such reason ready at hand if called upon to justify your implicit belief ascription at any time, but I sincerely doubt that this actually *is* what is operative in warranting your implicit belief ascriptions: is it the case, then, that your implicit belief ascriptions were not grounded until you took a particular philosophy course? It is even less plausible to suggest that this is what serves as a justifying reason for any given person performing intentional interpretation of others: this vastly overintellectualizes the process for the general public.

In considering these various distinct suggestions for what grounds the implicit belief ascriptions you have to make in interpreting someone else intentionally, it seems like the natural conclusion to draw is just that these implicit belief ascriptions—well, at least *some* of these implicit belief ascriptions—have no grounds at all.

Let's consider some objections to this conclusion.

3. Rethinking observation

It might be natural to feel, at this point, that the conclusion that some implicit belief

---

good support for *all* such generalizations. Another further problem is this: you would tend to notice, in trying to support these generalizations, more *confirmatory* instances than *disconfirmatory* instances (as is consistent with the literature on confirmation bias). Thus you would notice that people commenting on your outfit knew what you were wearing in looking at you, but you might never get that all your fellow subway riders are paying no attention whatsoever to your outfit despite their staring mindlessly in your precise direction.

ascriptions are groundless relies on a particular understanding of what is involved in observing others' behavior and interpreting it intentionally. There is one obvious sense in which this is true: this conclusion relies on the fact that intentional interpretation requires implicit belief ascription to other people—indeed, implicit ascription of *true* beliefs to other people.

Perhaps, then, it's the operative understanding of what is involved in such observation that matters to the conclusion—and perhaps some alternative understanding might allow us to see that we need not accept this requirement on intentional interpretation of others. In particular, you might feel that intentional interpretation has been made out to look extremely *theoretical* in nature, when in fact it involves *simulating* other people's minds. Or you might take it that intentional interpretation involves *analogical inference* from your own case to others that itself attaches intentional descriptions (contingently) to particular instances of observed behavior. Or you might just think that observation is richer than I have been making it out to be: you can *just see* that someone is tapping her pencil in order to annoy you, and your observations do not in the slightest underdetermine an intentional description of what someone else is doing.

I'll take each of these suggestions in turn. None of them helps.

3.1. Interpretation by simulation

Some cognitive scientists—including many philosophers—argue that we ascribe beliefs to others by running a *simulation* of how they are thinking and reasoning, rather than interpreting others by forming some sort of *theory* and reasoning with it as a scientist might.[187] Robert Gordon, for example, has argued that we can simulate, by allowing our regular rational capacities to work in an off-line way, what sort of decision procedures or reasoning processes another person is using at any time.[188] In observation of behavior, you might try running one such simulation to see if it results in the observed behavior—and thus tentatively conclude that the observed agent was thinking in the way that you just simulated, in a way that incorporates beliefs and desires and all sorts of mental states and feelings that you are seeking to ascribe in intentional interpretation.

This way of thinking about other-ascription of belief doesn't help us see how implicit ascriptions might have grounds precisely because it also *depends* on the interpreter's making some such implicit other-ascriptions of belief. To run a simulation in your own mind is to assume a similarity of background beliefs already; you make adjustments to pretend to have beliefs you yourself take to be *false* "only when necessary, only when something in the other's behavior doesn't fit."[189] To use this methodology is just to use what Gordon calls a "principle of least pretending," and given the 'principle of least pretending' mentioned earlier, "our belief attributions would be in accord with something like the 'principle of charity' put forward by Quine and Davidson: roughly, that one should prefer a translation that maximizes truth and rationality."[190]

If simulating in this way just involves making such implicit other-ascriptions of beliefs that you yourself take to be true, though, once again a regress threatens if we try

[187] Gopnik (1993).
[188] Gordon (1986).
[189] Gordon (1986), p.164.
[190] Gordon (1986), p.167.

to understand those implicit ascriptions as formed by way of simulation. Whether or not this proposal is true, it does not offer any way of understanding the implicit belief ascriptions in question as anything other than groundless.

3.2. Interpretation by analogical inference

Would it help to think of other-ascriptions of belief along the lines of *analogical inference*? This idea, famously proposed by Russell, takes it that ascribing a particular mental state to another person involves applying causal knowledge gained in observation of your own case.[191] Prima facie, this suggestion has a particular advantage over the others just considered. The way in which you build causal knowledge, according to Russell, is by recognition of the causal connections between your *own* thoughts and behavior. But when you *self*-ascribe beliefs, as part of recognizing such causal connections, you do not run up against the same underdetermination problem that you face in trying to come to intentional descriptions of *others*. You yourself have in hand, necessarily, the correct intentional interpretation of your own beliefs and actions. This is what Davidson means when he denies that there can be any question of interpreting *your own* attitudes. "Knowledge of the contents of our own minds must," according to him, "in most cases, be trivial. The reason is that, apart from special cases, the problem of interpretation cannot arise. When I am asked about the propositional contents of my mind, I must use my own sentences."[192] This point is deeply connected with Anscombe's claim that performing an action that is intentional under a given description is to have awareness of what you are doing under that description.[193] In both cases, the intentional description of some attitude or behavior of *yours* is immediately available to you in virtue of that attitude's *being yours*.

This may seem helpful in solving the underdetermination problem, because we need not rely on uncertain observational inferences from *others'* behavior in order to build a causal picture of how certain beliefs produce certain kinds of behavior. And if you have that causal picture in hand, it seems, you can then simply apply it to observed behavior in order to determine the correct intentional description of the behavior in question. There is a slate of familiar worries for the view that analogical inference is the way in which we know other minds. Concerns have been raised about the poverty of the causal data you gain in analyzing your own behavior, the illegitimacy of generalizing from your own case to others', and the failure of analogical inference as a response to *skepticism* about other minds.[194]

Even leaving aside these familiar worries, though, there is a special problem for understanding analogical inference as the way in which we ascribe beliefs to other people whose behavior we observe. It is, not surprisingly, another underdetermination problem that affects the way we *apply* the causal knowledge we take ourselves to have about which beliefs cause which kinds of behavior. The problem has two aspects. First, the pieces of causal knowledge we can apply to understand others' behavior relate *intentionally described* behavior to intentional mental states; and second, the causal

[191] Russell (1948).
[192] Davidson (1991/2001), p.217.
[193] Anscombe (1957).
[194] Gordon (1986) pp.159-60, Hyslop (2014), Malcolm (1958).

inferences are only valid given *ceteris parabus* clauses that build in *other* intentional mental states.

On the first point: consider seeing someone kick a ball across a field, and trying to determine her mental states by causal analogical inference. In your own case, you can meaningfully understand *passing the ball* as caused by a belief that *your teammate is over there*. And you can meaningfully understand *getting the ball out of your way* as caused by a belief that *you might trip on it while you're running*. If the ball kick doesn't wear its intentional description on its sleeve, which of these available causal inferences should you make? You would need to have an intentional description in hand already in order to apply the causal knowledge. But an intentional description is just what you are seeking.

On the second point: you know (if your causal inferences have any sensitivity whatsoever) that a belief that *your teammate is over there* only causes *passing the ball* when you also have a desire *to give your teammate the ball*. And you know that *getting the ball out of your way* is only caused by the belief that *you might trip* only when you also have the desire *not to trip*. This point is essentially the same point as the point that intentional intepretation requires implicit belief ascription: you need to have in hand something about an observed person's state of mind in order to apply the correct intentional description to her behavior. But again, this is just what you are seeking.

In short, to know how to apply the causal knowledge gained by self-analysis, you need first to settle, for some piece of behavior, what its subject takes it to be, and what other intentional attitudes that subject has. But settling those questions already involves implicit ascription of beliefs to the subject being analyzed. But that is precisely what we wanted to use analogical inference *to avoid doing*. Unfortunately, it seems, analogical inference fares no better with respect to avoiding groundless implicit ascriptions of beliefs to others.

3.3. Enriched perception

There may be an altogether different problem with the way that we have been understanding observation in this paper. I argued above that observation of behavior *underdetermines* its intentional description. But it seems to be a plain fact that you can *see* what someone is doing, in intensional terms, and thereby actually know what it is that she is doing, in the very terms in which she understands it herself. I can *see* that my aunt is regretting taking me on this camping trip, as she throws all of our stuff back into the car and glares at me; I don't need to *do* anything extra in order to find that out. I can *see* that a familiar colleague is about to raise the same objection that she has raised every time we've talked about perceptual illusions before, even before she opens her mouth to speak.[195] These kinds of examples are completely ordinary, and these ways of expressing intentional interpretation are deeply familiar.

Now, if observation of other people is really as rich as all that, and you really can *see* what someone is thinking or feeling, shouldn't we reject the claim that observation of behavior underdetermines its intentional description? If we reject that claim, then the arguments of the previous section cannot succeed. That would mean that no implicit

[195] Of course, these points apply not only to the visual modality; you can also *hear* in a sigh that someone is stressed out about the business's finances, or *feel* in the way someone grips your arm that he thinks the rollercoaster is going way too fast.

belief ascription is groundless after all. Indeed, it might imply that the grounds are themselves perceptual or observational, and that there's no more to say about it.

I don't think that is the right conclusion to draw from the undeniable fact that sometimes you can just *see* what someone is thinking. On the contrary: the fact that you can just *see* what someone is thinking *raises* the same questions that we have been discussing here. What is involved in *seeing* that someone is thinking something, or feeling a certain way, or doing something on purpose? Part of what is involved is making implicit belief ascriptions. To return to the example of the irritating SAT-taker: to *see* that she is trying to annoy you is also to ascribe to her, implicitly, the belief that the sound of the pencil tapping will probably annoy you. To *see* that she is trying to signal the answers to Garth is also to ascribe to her, implicitly, the belief that Garth can hear her tapping.

Compare conceptual enrichment of perception in other areas. The fact that you can *just see* that something is a pine tree does not mean that all that is involved in seeing *that* is just opening your eyes and letting the world impinge on your visual cortex. We can ask, in a straightforward way, what is involved in seeing something as a pine tree, and ask what grounds or (more broadly) warrants any assumptions that go into seeing the tree as a pine tree. That is just what we are doing here for intentional interpretation.

Unfortunately, understanding our perceptual observations as sufficiently conceptually enriched to encompass intentional interpretation—plausible as it is—will not forestall the lines of questioning that we have followed here, and it certainly doesn't provide any other way of understanding the *grounds* on which you make implicit belief ascriptions.

The general strategy that involves rethinking what is involved in our intentional interpretation of other people's behavior has not, thus far, offered any hope of avoiding the conclusion that some of our ascriptions of beliefs to other people—and in particular, some ascriptions of beliefs that we as interpreters take to be *true*—must be made without any *justifying reasons* at all.

4. Rethinking belief

It may also be objected that we have been working with a particular view of belief in mind, and that another view of belief would allow us to avoid the conclusions I have drawn above. It is not obvious to me that what has been argued here depends on any particular one view of belief. But just to clarify, let's consider some positions about belief that may seem, prima facie, to offer us a way of escaping one of the premises of the argument above: dispositionalism, functionalism, interpretationism, and instrumentalism.[196] None of these views offers any escape from either the claim that we *must* make implicit belief ascriptions in intentional interpretation of observed behavior or the claim that some such ascriptions are entirely groundless.

4.1. Dispositionalism and functionalism

It may seem possible to avoid the conclusion that some implicit belief ascriptions are

[196] Of course this is not an exhaustive list of the metaphysical views you could have about belief— but I take it that these views are the views that most naturally seem to offer any escape from the conclusions of this paper by making belief metaphysically *thinner* and thus (one might think) easier to attribute on the basis of observed behavior.

made *groundlessly* by thinking of belief as essentially a dispositional (and partly behavioral-dispositional) state, or by thinking of belief as that which plays a particular functional role in producing and responding to behavior. These positions are *dispositionalism* and *functionalism* about belief. They are, of course, not the same view of belief—but here I will treat them together, as they both may seem to offer help in a particular kind of way, and they both fail to do so in the same kind of way.

If belief that *p* is defined in particular terms that make explicit reference to behavior in extensional terms, then perhaps we need not worry about the underdetermination problems that we recognized above.[197] For observation of behavior just is sufficient to *determine* the existence of a particular disposition, extensionally described, and certainly sufficient to *determine* the presence of some behavior that might be caused by or cause some belief as functionally defined. And determining that much might just be enough to determine a belief, *intensionally* described.

The problem, of course, is that while such observation might be sufficient to determine some such disposition or the existence of some such behavior, on any *adequate* dispositionalist or functionalist theory belief that *p* will turn out to be an incredibly complex state, associated with all sorts of defeasible dispositions and behaviors in distinct circumstances. Any behavior or behavioral disposition that you can establish by pure observation will also be associated, definitionally, with many *other* beliefs, as they combine with distinct desires or intentions. Thus to establish that an observed agent has some disposition or is doing some particular thing (extensionally described) is not yet to settle *which* of the many and various beliefs associated with such behavior to attribute. And this is just the kind of underdetermination we have been discussing all along.

It is not clear, either, how endorsing dispositionalism or functionalism about belief could ever offer us any other options for understanding the *grounds* on which some implicit belief ascriptions are made. Given that fact, and given the same underdetermination problem discussed above, it's clear that neither view will allow us to avoid the conclusion that some implicit belief ascriptions are made *groundlessly*.

4.2. Interpretationism

Perhaps relating belief more directly to our interpretation of others will allow us to avoid the underdetermination problem. Consider *interpretationism*, a view advanced by Daniel Dennett, on which "*what it is* to be a true believer is to be an *intentional system*, a system whose behavior is reliably and voluminously predictable via the intentional strategy."[198] On this position, people genuinely have beliefs, but the presence of those beliefs is *definitionally* linked to the practice of interpretation.[199]

Unfortunately, this view fares no better on avoiding the underdetermination problem.

---

[197] Of course, some versions of dispositionalism or functionalism about belief will not define the relevant behavior in purely extensional terms. These versions of the view will face additional problems of the same kind that faced the suggestion that we interpret others on the basis of *analogical inference*: even trying to apply some such view requires intentional interpretation.

[198] Dennett (1981/2001), p.557.

[199] Schwitzgebel (2015) also claims that Davidson himself is an interpretationist about belief. I don't see the case for Schwitzgebel's attribution of this view to Davidson; I think Davidson's position is more complex and less easy to categorize than Schwitzgebel.

Attributing beliefs to someone by observation may itself be ratified as a practice when you have appreciable success in predicting their future behavior, but interpretationism (at least as presented in Dennett's 1981 paper) does not specify any way of ratifying, by prediction, that you have attributed the *right* beliefs to another person. Indeed, several sets of belief ascriptions might do the job just as well.

Even given some such way to ratify your implicit belief ascriptions, you would need a place to get started on the project of intentional interpretation before you could even think about ratification of the particular ascriptions you are making. For this reason, Dennett recognizes the importance of using the principle of charity even on an interpretationist view. He writes: "one rule for attributing beliefs in the intentional strategy is this: attribute as beliefs all the truths relevant to the system's interests (or desires) that the system's experience to date has made available... an implication of the intentional strategy, then, is that true believers mainly believe truths."[200] It is this principle, not any extra *reason* to attribute some belief rather than another belief, that allows us to use the intentional strategy at all.

There are, on interpretationism, still *distinct* ways of interpreting someone, such that distinct implicit belief ascriptions are not collapsible into one and the same based on the predictions that they make. (More on that idea under "instrumentalism.") If that is still the case, nothing in interpretationism has any hope of eliminating the problem of underdetermination. And nothing in this view offers any extra *grounds* for implicit belief ascriptions. This view cannot avoid the conclusion that some implicit belief ascriptions are made—indeed, *need* to be made—groundlessly.

4.3. Instrumentalism

In later writing Dennett seemed to espouse a variety of *instrumentalism* about belief, which is more committal than his previous *interpretationism* on the matter of what it is to have some particular belief, rather than what it is to be a "true believer" in general.[201] What it is for a subject to have a particular belief is for there to be a reliable pattern into which that belief figures such that the subject's behavior can be effectively predicted by recognizing that pattern. Dennett clarifies that there may be no fact of the matter, when two patterns produce equally reliable predictions of behavior, which of the two involves the "correct" belief ascriptions, or which of the two captures the "reality" of the believer's total doxastic set. It is hard to tell from the armchair how often two appreciably distinct total systems will produce appreciably distinct predictions about behavior, but we can nonetheless consider these cases in the abstract.

Note that part of the way that instrumentalism circumscribes the possibility of underdetermination is by reference to prediction. This places constraints on the implicit belief ascriptions with reference to further unobserved behavior, where that further behavior *once observed* will help constrain the class of reasonable belief ascriptions. This sort of constraint might at first look like an advantage of instrumentalism, but it is important to remember that on any serious account of intentional interpretation, more observation will further constrain available theories of the total intentional state of an

[200] Dennett (1981/2001), pp.557-9.

[201] Dennett (1991). I follow Schwitzgebel (2015) in calling this view "instrumentalism," which is not to deny that this view is also a *realist* view, as Dennett (1991) himself points out (p.51).

individual. It is particularly what instrumentalism says about the *nature* of belief as connected with prediction that sets it apart from other theories; it does not, in fact, have in hand an extra constraint that would help in intentional interpretation. After all, if you are looking at your neighbor in the SAT and trying to figure out what she is doing, you don't yet have future observation to go on: you are trying to come up with an intentional description *now*, which description may or may not prove to be realistic later on.

Still, perhaps there is a way of understanding groundless implicit belief ascriptions as especially suited to an instrumentalist theory of the nature of belief. If there's nothing more to having a belief than being efficiently and effectively predictable according to belief ascriptions—and here I am abstracting away, somewhat, from Dennett's (1991) particular brand of instrumentalism—then perhaps the ascription of true belief to others is particularly well warranted, given that it allows for easy and efficient interpretation. What easier way to remember what someone else believes than to attribute to them as many true beliefs as is possible, given the way she or he behaves? Perhaps, one might suggest, following the principle of charity is a strategy that you can try a couple of times, and, at least after seeing how it produces efficient and effective predictions, you can have reason to use in the future.

Note that this suggestion subtly transmutes *pragmatic* reason into *epistemic* reason. Ordinarily, a pragmatic reason to do something—even to ascribe a belief—does not imply any particular epistemic reason to do the same thing. But one could argue that the advantage of instrumentalism lies in its identification of the two things: given the facts about what it is for a belief ascription to be true of someone, a pragmatic reason is an epistemic reason. (This is, of course, a very rough way of putting a point that deserves more sophisticated formulation, but it should be good enough for current purposes.)

The problem with this suggestion is that it seems you would have to recognize the instrumentalist nature of belief in order to take your own pragmatic reasons as epistemic reasons—and that is what's needed for you to have *grounds* for implicit belief ascriptions. (Recall the comment on internalism and grounds made earlier.) But it is highly unlikely that we all think of belief ascriptions as instrumentalist, and even less likely that our implicit belief ascriptions could be *grounded* only once we recognized the truth of this philosophical theory. This seems like an overintellectualization of the way that any of us chooses implicit belief ascriptions to make in intentional interpretation. I think a more fitting thing for the instrumentalist to say is that implicit belief ascriptions, such as they are, do not actually need epistemic grounds whatsoever. That is perfectly consistent with their sometimes being groundless.

How about a yet stronger version of instrumentalism, on which two total intentional descriptions that make all the same predictions actually say the same thing about a person—despite, perhaps, apparent direct disagreement in implicit belief ascriptions? Could this stronger version of instrumentalism (stronger, that is, than Dennett's 1991 version) save us from the problem of underdetermination altogether?

It doesn't seem that it could, for the reason that we are concerned in this paper with *particular* belief ascriptions. The level on which the instrumentalist can identify two theories as actually being one and the same is the level of an intentional description of a total system (a person as a whole), and here we are asking about the grounds for particular belief ascriptions. It's not clear how the instrumentalist identification of two total description of an intentional systems could imply that any particular belief ascription

contributing to one such total description has particular grounds. As above, even this stronger instrumentalist position is, I think, best combined with a view that some implicit belief ascriptions need no grounds.

5. Consequences

The asymmetry between the first and the third person relevant to explaining first-person authority is not about groundlessness. In fact, the only asymmetry needed to explain first-person authority with respect to beliefs is an asymmetry in strength of warrant. Transparent self-attribution of beliefs (or intentions) grants especially secure knowledge that is not available from the third-personal perspective.

The more fundamental point is that groundlessness is at the source of *all* intentional descriptions of behavior observed third-personally. Above, I argued that *some* implicit belief ascriptions must be groundless, by way of eliminating potential grounds for such ascriptions. But if the argument stands, then *all* intentional description of observed behavior will depend upon *some* such implicit belief ascriptions, if perhaps only indirectly, by way of depending on *past* intentional descriptions of other observed behavior on the part of the same individual to be described.[202]

What this implies is that understanding observed behavior as expressing, determining, or fulfilling intentional attitudes on the part of the behaving subject involves making a certain kind of leap of faith, at least on some implicit level. All our understanding of the people we see and hear as minded beings with intentional stances on the world depends on certain groundless ascriptions of true beliefs to them. What I hope this paper brings out is the fundamentality of the general practice of groundless true belief ascription to others. As Dennett powerfully puts the point, without intentional interpretation, "human activity would be just so much Brownian motion; we would be baffling ciphers to each other ... we could not even conceptualize our own flailings."[203] And if intentional interpretation, as I have argued, requires implicit but groundless ascription of true beliefs, then to *be people* to one another—to see others as people, and to be seen as people by others—we need to assume, with no reason, they have a lot of true beliefs.

Those are the main theses of this paper. In the remainder of this section, I'd like to speculate a little bit about other potential consequences of these claims. In particular, I'll first consider some of the claims that Davidson has made about the impossibility of error in using the principle of charity. Then I'll ask whether these theses straightforwardly open up any epistemological problem, and tentatively conclude that they do not.

---

[202] Does this apply even to speech behavior in a language that you (the interpreter) understand, where said speech explicitly identifies an intentional description of the person to be interpreted? Imagine someone tells you "I believe Trump will not get the Republican nomination." *Even then*, to apply the intentional description to this person that *she believes Trump will not get the Republican nomination*, you have to assume she has certain true beliefs related to her communication: for instance, that you will understand her words, that you will take her to be truthful, and so forth. This is the point that I have taken Davidson to overlook in not extending his thoughts on radical interpretation to everyday situations.

[203] Dennett (1991), p.29.

5.1. Davidson on impossibility of error

Davidson, in recognizing the indispensability of the principle of charity in *radical* interpretation, has suggested that the practice of ascribing true beliefs to others in order to interpret their behavior is beyond criticism. Its necessity, he implies, is actually what provides the practice as a whole with some justification: "what justifies the procedure is the fact that disagreement and agreement alike are intelligible only against a background of massive agreement."[204] He writes that "charity is not an option, but a condition of having a workable theory," and so "it is *meaningless* to suggest that we might fall into massive error by endorsing it."[205]

I am not sure that this is right; I don't understand why it must be right, unless we commit to a certain form of verificationism about meaning on which only falsifiable statements have any meaning. Moreover, as Barry Stroud has compellingly argued, the fact that this practice as a whole is indispensable to our interpretive practices in general does not yet imply that it is justified—let alone justified in an epistemic, rather than a pragmatic or ethical, way.[206] I'll leave aside this broad question for now, but here I'd like to note that *even if* it were meaningless to suppose that the practice could lead to "massive error," and *even if* the practice were justified in a distinctively epistemic way, the particular implicit belief ascriptions that it produces may still remain entirely groundless. Even if following the principle of charity is generally justified, the principle itself does not directly determine how strictly to apply it, or how to adjust our intentional descriptions of others based on behavior that seems to imply the presence of *false* beliefs.

5.2. Resisting skepticism

We might be tempted to use the claim that some implicit belief ascriptions are groundless—and the further claim that all intentional interpretation of observed behavior must involve some such implicit belief ascriptions—that our knowledge of others under intensional descriptions is under threat, or requires further explanation. We may even be tempted towards skepticism about knowledge of other minds of a new sort—not the kind that doubts whether other minds exist, or whether they are conscious, and so forth, but rather doubts *what we know* about other minds to the extent that other people are reduced to inscrutable puzzles in the face of our attempts at intentional interpretation.

These conclusions are not mandatory given what I have said in this chapter. It is hard to assess here. To be rigorous about the epistemological implications of the groundlessness of implicit belief ascriptions, we would need to understand better exactly *how* these belief ascriptions are 'involved in' intentional descriptions of observed behavior. Do they themselves play a direct justificatory role for the target explicit intentional description? Do they *mediate* whether certain observations can be taken as evidence for the target description? Are they merely *implied by* target intentional descriptions? Perhaps some of the issues about how we *actually* go about providing intentional descriptions in observing people's behavior discussed above—the points on

[204] Davidson (1973/2001), p.137.
[205] Davidson (1974b/2001), p.197, emphasis added.
[206] Stroud (1984).

which Gopnik, Gordon, and Russell disagree—matter more here.[207]

Yet even without settling these points, it seems there are ways of making sense of intentional knowledge of other people *as* knowledge. We can understand implicit belief ascriptions as having associated entitlements, even if they lack justificatory grounds. Or we can understand intentional knowledge as depending on ultimately warrantless implicit belief ascriptions, but explain—perhaps by an argument by analogy, comparing such knowledge to knowledge in other domains—that that is not itself a problem. Or we can understand the groundless belief ascriptions as warranted by way of familiar externalist routes.[208]

It is worth taking a moment to diagnose some of the epistemological discomfort we might feel in accepting that some implicit belief ascriptions are made groundlessly. Whenever you ascribe a belief to another person, you must respond to any demands for *grounds* for your belief ascription. For example: if you observe someone's behavior and conclude that she believes it will rain, you should be prepared to explain your reasons. Others can challenge you on your ascription, bring to bear further evidence one way or the other, and even sometimes move you to retract or modify your ascription. And it seems that we very rarely—or perhaps never—do actually make belief ascriptions to others that we cannot justify with respect to at least *some* (defeasible) grounds.

But while this is a genuine and natural source of discomfort, it should not be taken as a reason to think that we do *not* make *implicit* belief ascriptions groundlessly. It is consistent to think that we must rely on a good deal of groundless ascription in order to have any knowledge of others' intentional stances at all by observation, while also recognizing that each such ascription is reviewable, defeasible, and fallible—especially when brought to the level of explicit consideration. Any *particular* belief ascription is also ultimately dispensable to the practice of intentional interpretation, while the general practice of making such implicit belief ascriptions—and groundlessly doing so—is not.

Conclusion

Groundless ascription of belief is essential to understanding both ourselves and others as minded creatures with intentional stances on the world—creatures with beliefs, desires, and intentions, who can perform actions for particular reasons and in so doing understand what they are doing in particular ways. Recognizing this important commonality between the first- and third-personal perspective helps us understand what it is to understand ourselves and other people in this way on a fundamental level. As Davidson put it, "Charity is forced on us; whether we like it or not, if we want to understand others, we must count them right in most matters."[209] I would add: we have to do that *groundlessly*.

---

[207] Gopnik (1993), Gordon (1986), Russell (1948).

[208] This last option does not necessarily involve abandoning *all* internalist constraints on warrant. In particular: you might endorse certain constraints on complexes of implicit belief ascriptions, like a rationality constraint, that would make the person to be interpreted come out to be broadly rational. Davidson seems to endorse these constraints as well: we don't *only* understand others as believing true things, but also as rational believers who are minimally inconsistent. Whether or not these two constraints are truly independent is an interesting question in its own right. See Lewis (1974) for discussion of this point.

[209] Davidson (1974b/2001), p.197.

Chapter 6. The Value of Transparent Self-Knowledge

The discussion of this dissertation has focused on transparent self-knowledge. There are various other forms of self-knowledge that have not been addressed in any of the previous chapters. These include: self-knowledge about your bodily states and your health; self-knowledge about your strengths and weaknesses; self-knowledge about your position in the social world; self-knowledge about your personal character; and many more kinds of self-knowledge besides these.

When we contrast transparent self-knowledge with these other kinds of self-knowledge, it is easy to feel that transparent self-knowledge is not particularly valuable to each of us in our personal lives. That is partly because these other kinds of self-knowledge are clearly important to gain. It is important to know your allergies, for example, to avoid illness or accidental death. It is good to know your intellectual biases so you can work to correct them. Knowing your social rank will help you avoid faux pas. Knowledge of your tenacity or bravery will help you face the future with optimism.

It is also easy to feel that transparent self-knowledge cannot hold much personal importance because it is guaranteed for each of us. We value those kinds of self-knowledge that are difficult to gain—the kinds that we might fail to have. The philosophers might find abstruse interest in transparent self-knowledge to illuminate the nature of belief, intention, and the first-personal perspective, but surely each of us as private individuals would be daft to pursue this kind of self-knowledge.

This is a common thought, well expressed by Quassim Cassam.[210] It contrasts the 'ease' or simplicity of transparent self-knowledge with the difficulty and substance of non-transparent self-knowledge.

The point is well motivated. Non-transparent self-knowledge really is difficult to gain in many important cases, and often valuable. It is also true that trivial or guaranteed self-knowledge does not make sense to pursue in your personal life. However, neither of these points implies that no transparent self-knowledge is valuable or difficult to gain.

In this chapter I show that transparent self-knowledge of your *diachronic* states is something that we value highly in our personal lives. Even though transparent self-knowledge of your synchronic mental acts and states is guaranteed, transparent self-knowledge of your diachronic states is not. It makes sense, then, to pursue this type of self-knowledge—diachronic transparent self-knowledge. Pursuing it, however, is not a distinctively epistemic task. To gain better diachronic transparent self-knowledge is to constitute yourself so as to be more readily knowable in transparent ways.

Here is the structure of this chapter. In Section 1 I argue that failures of diachronic transparent self-knowledge arise from inconsistency or inconstancy in your attitudes. While inconsistency is clearly to be avoided, it is not clear why inconstancy should be. In Section 2 I use examples of praise and criticism in literature to demonstrate that we genuinely do value constancy and disvalue inconstancy. In Section 3 I explain these expressions in a way that brings transparent self-knowledge into a central position in these forms of praise and criticism. Section 4 asks whether constancy really is valuable.

[210] See Cassam (2014).

1. Limits and failures of transparency

In Chapter 2 I showed that transparent self-attribution of belief is infallible in cases of synchronic self-knowledge, and in Chapter 4 I showed the same for transparent self-attribution of intention. When you transparently self-attribute one of these attitudes, you cannot fail to gain knowledge of what you believe or intend at that very moment.

However, once we consider the full set of beliefs and intentions that an agent can be truly said to have, we can see that transparent self-knowledge does not constitute infallible and complete self-knowledge of all that an agent believes and intends.

There are two ways in which transparent self-attribution is imperfect in capturing an agent's total set of beliefs and intentions. The first has to do with beliefs and intentions that are hidden to the agent's conscious consideration, and the second has to do with beliefs and intentions that change over time. I'll consider each in turn.

1.1. Inconsistency and hidden attitudes

By definition, transparent self-attribution of a belief or an intention involves making a conscious and intentional mental act with the same content as the belief or the intention to be self-attributed. When you transparently self-attribute a belief that *p*, you judge that *p*. When you transparently self-attribute an intention to Φ, you decide to Φ.

However, it is possible for you to have beliefs and intentions that are 'hidden' to your conscious consideration. You can have beliefs whose contents you would never reaffirm in conscious judgment. You can also have intentions whose plans you would never reaffirm in conscious decision.[211] These attitudes are attitudes that you simply cannot self-attribute transparently. If you tried to transparently self-attribute a belief whether *p* or an intention concerning Φ-ing, your attempt to do so would either result in a failure or a conflicting self-attribution. You might erroneously think that you had no belief or intention on the relevant matter, or you might actually self-attribute a belief that it's *not* the case that *p*, or an intention *not* to Φ.

Note that these possibilities are perfectly consistent with the synchronic infallibility of the transparency method. The transparency method (for either belief or intention) by definition involves a positive attribution of some belief or intention, so a *failure* to use the transparency method that results in the belief that you have no belief or intention on the relevant matter is not a way in which the *transparency* method delivers erroneous belief. Moreover, to the extent that you self-attribute a belief or intention that directly conflicts with a 'hidden' attitude you have but would never consciously re-affirm, your transparent self-attribution is still *true*: your deciding to Φ or judging that *p* consciously ensures that you do have the relevant intention or belief *as well* as the hidden one.

Nonetheless, even though your transparent self-attribution (if successful) would deliver a true verdict about *something* you believe or intend in either situation, neither situation is one that you would want to be in. To have two inconsistent beliefs or two directly contrary intentions is to be fragmented and irrational in your stance on the world. Having hidden beliefs or intentions is what allows for this kind of synchronic and thus

[211] This point is somewhat more controversial than the same point about belief; see, e.g., Wallace (2002), p.22.

irrational fragmentation.[212] Even if you do not suffer from this kind of inconsistency, and your belief or intention is merely hidden from you, a hidden attitude of this kind is not playing the total role it *should* play in guiding your conscious thought. A belief that *p* *should* lead you to think that *p* when you consciously consider the question. And an intention to Φ *should* lead you to plan around Φ-ing when you consider whether to Φ. Moreover, a belief or an intention that cannot be raised to consciousness is one that is improperly shielded from the careful, effortful, and delicate work of conscious reconsideration. Your best conscious reasoning about what is true, or about what to do, should be able to dislodge beliefs or intentions you already have. Hidden beliefs and intentions cannot be dislodged in this way. They can persist when they should not.[213]

This gives us a sense of the value of transparent self-knowledge—or, rather, a sense of the value of **complete** transparent self-knowledge. If your transparent self-knowledge is *complete*, you have no attitudes that are hidden to you in conscious consideration. Any belief that you have is one that you would reaffirm in conscious judgment, and any intention you have is one that you would reaffirm in conscious decision. That is just what is needed for transparent self-knowledge to be complete in this sense.

1.2. Inconstancy and change of attitude

Transparent self-knowledge, then, can fail to be complete in the sense that it can fail to capture *all* that you believe or intend at one moment. To have complete transparent self-knowledge is also to be free of hidden attitudes, which are themselves problematic for the reasons just described.

There is also another limitation on transparent self-knowledge. Transparent self-knowledge is infallible in the moment, but it can fail to capture your states over time. A belief or an intention that you transparently self-attribute in one moment might be merely temporary. It could fail to persist into the future, or it could be brand new. Neither fact about the diachronic profile of the attitude you self-attribute in this way is itself captured by the transparency method. The transparency method is silent on diachronic changes in your attitudes—even though you can, and do, use the transparency method to self-attribute beliefs and intentions that *do* persist over time.

Whether or not your transparent self-attribution of a belief or intention actually constitutes **diachronic** self-knowledge is not a matter of performing the transparency method well. It is a matter of whether the belief or intention you self-attribute in this way actually persists through time. It depends on the **constancy** of the attitudes themselves. If your beliefs or intentions on some matter are **inconstant**, a momentary transparent self-attribution of *some* belief or intention will not constitute diachronic self-knowledge. To have constancy in your attitudes is to maintain them over time. Note that inconstancy is not a matter of *not* making up your mind: it's a matter of *changing* your mind.

This fact suggests that there is another way for transparent self-knowledge of a certain kind to be valuable. Insofar as you have *diachronic* transparent self-knowledge, your attitudes are also constant. This formulation mirrors the formulation from the previous subsection: insofar as you have *complete* transparent self-knowledge, your

[212] Recall the arguments from Chapters 3 and 4 on the conditions under which someone can have inconsistent beliefs or directly contrary intentions.
[213] Compare Moran (2001) on estrangement.

attitudes are not hidden to conscious consideration.

It was a simple matter to recognize the value of lacking hidden attitudes. However, it is as easy to see the value of having *constant* attitudes. Sometimes it is good for your beliefs and intentions to change. When you gain new evidence, or change what you value, your beliefs and intentions should adjust accordingly. Making these adjustments in a changing world is part of taking responsibility for your beliefs and your intentions. Why, then, is it valuable to have diachronic transparent self-knowledge, if all that implies is that your attitudes themselves are constant over time?

Answering this question will be the focus of the remainder of this paper. I will begin in the next section by showing that we *do* value constancy in our own and others' attitudes. In fact, we criticize inconstancy, and praise constancy, by way of making claims about self-knowledge.

2. Constancy in literature

Although it is difficult to recognize the value of constancy in belief and intention from a philosophical perspective, we often praise constancy and criticize inconstancy in ordinary situations. Some such examples of praise and criticism arise in famous literary contexts. In this section I will give three examples of such praise and criticism.

Importantly for our purposes, this praise and criticism is doled out in terms of self-knowledge. Constancy can be praised by praising someone's self-knowledge, and inconstancy can be criticized by criticizing someone's lack of self-knowledge.

Let's begin with an example from Shakespeare's *Tragedy of King Lear*. In the first scene of the play, the aging Lear calls a meeting with his three daughters to determine their inheritance. He asks for a profession of love from each to claim her share of his kingdom. Goneril and Regan make effusive toasts, but Cordelia—who really loves him best, and who has long been his clear favorite—finds herself speechless. Her genuine fondness for her father is embarrassed and silenced by the demand for a show. In a rage, Lear disowns and disinherits Cordelia. When his most trusted adviser, the Earl of Kent, cautions him against this terrible act, Lear turns his wrath on Kent: he gets ten days to flee the kingdom before he is pursued and killed.

This is a clear show of inconstancy on Lear's part. Where he previously favored Cordelia, and had intended to give her the best part of his holdings, he now disowns her. He had previously believed she loved him best, and now he changes his mind.

After the main drama of the scene is over, Goneril and Regan lament Lear's behavior:

GONERIL You see how full of changes his age is. The observation we have made of it hath been little. He always loved our sister most, and with what poor judgement he hath now cast her off appears too grossly.

REGAN 'Tis the infirmity of his age; yet he hath ever but slenderly known himself.

GONERIL The best and soundest of his time hath been but rash; then must we look from his age to receive not alone the imperfections of long-engrafted condition, but therewithal the unruly waywardness that infirm and choleric years bring with them.

REGAN Such unconstant starts are we like to have from him as this of Kent's

banishment.[214]

Here Lear's shocking inconstancy, expressed in terms of his "waywardness" and his "inconstant starts," is criticized by way of a claim about self-knowledge. Regan's claim that her father "hath ever but slenderly known himself" is clearly something negative—something to be lamented or even to be censured. The claim about Lear's lack of self-knowledge is, in context, a way of emphasizing his lack of consistency over time. The only question that Goneril and Regan are debating is whether Lear has always been so inconstant (one of the "imperfections of long-engrafted condition"), or whether his inconstancy is a matter of the "infirmity of his [old] age."

The same kind of criticism is made in Oscar Wilde's play "A Woman of No Importance." Gerald Arbuthnot is all but ready to take an appointment with Lord Illingworth—whom Gerald does not know is his own father—when his mother urges him against it in private. Since Mrs. Arbuthnot had just approved the move in conversation in public, Gerald is filled with consternation. He upbraids his mother in private:

> Mother, how changeable you are! You don't seem to know your own mind for a single moment. An hour and a half ago in the Drawing-room you agreed to the whole thing; now you turn round and make objections, and try to force me to give up my one chance in life.[215]

Note the similarity between the contexts of these criticisms. Lear has suddenly turned against those he has long loved best. Mrs. Arbuthnot seems (to Gerald) to have changed her mind at a moment's notice. It is this inconstancy that makes Gerald's criticism about his mother's lack of self-knowledge apt. His complaint that she doesn't "seem to know [her] own mind for a single moment" is a pressing restatement of the complaint about her changeability. Once again inconstancy is criticized in terms of a lack of self-knowledge.

The constancy that Lear and Mrs. Arbuthnot lack is treated as something to be pursued or cultivated. It is a quality that Jane Austen, in classic Austenian free indirect style, imputes to a young man called Edmund Bertram in *Mansfield Park*:

> Edmund was at this time particularly full of cares: his mind being deeply occupied in the consideration of two important events now at hand, which were to fix his fate in life—ordination and matrimony—events of such a serious character as to make the ball, which would be very quickly followed by one of them, appear of less moment in his eyes than in those of any other person in the house. On the 23rd he was going to a friend near Peterborough, in the same situation as himself, and they were to receive ordination in the course of the Christmas week. Half his destiny would then be determined, but the other half might not be so very smoothly wooed. His duties would be established, but the wife who was to share, and animate, and reward those duties, might yet be unattainable. He knew his own mind, but he was not always perfectly assured of knowing Miss Crawford's.[216]

---

[214] Shakespeare (2005), I.i.288-300, emphasis mine.
[215] Wilde (1893), emphasis mine.
[216] Austen (1814), emphasis mine. "Free indirect style" is a term coined by James Wood to capture the stylistic strategy of incorporating a character's personal attitudes towards states of

This example involves a commendation of self-knowledge that also appears to be a commendation of constancy. Edmund's fixation on his goals of ordination and matrimony, and his understanding of those goals as constituting his "destiny," speaks to the constancy of his intentions. The praise here associates constancy with self-knowledge just as the previous examples associated inconstancy with its lack.

Constancy and its lack was discussed at length by the sixteenth-century French essayist Michel de Montaigne, most notably in his essay "On the inconstancy of our actions."[217] He complained about a pandemic of 'inconstance':

> vacillation seems to me to be the most common and blatant defect of our nature … Of Man I can believe nothing less easily than invariability: nothing more easily than variability. … anyone who studies himself attentively finds in himself and in his very judgement this whirring about and this discordancy.[218]

Such vacillation or variability is the opposite of constancy. According to Montaigne, it is one of the gravest and most common vices.

These quotations from plays, novels, and essays establish that we do praise constancy and criticize inconstancy. They also manifest an association between constancy and self-knowledge. This association should be familiar from other everyday expressions. We can say "he knows what he thinks," "he knows what he believes," and "he knows what he intends to do," and each of these—in the right context, if not always—can be used to make a point about the subject's enviable, or impressive, constancy. The same point goes for phrases like "he knows his values," or "he knows his will." These can be made more specific as well. We can say "he knows what he thinks about North Korea," or "he knows what he wants to do about poverty," or "he knows his mind when it comes to coffee."

It's also worth noting that this sort of expression admits of comparative dimensions: we can draw attention to your greater constancy by saying *you* know your own mind better than I know mine. And the direct objects of the knowledge are themselves flexible too: to say that you have the sort of constancy at issue here, I can say that you "know your mind," "know yourself," or "know who you are."

To recognize this association in our everyday speech, however, is not yet to explain these expressions in full. What exactly are we saying of someone when we say she doesn't know her own mind, or she doesn't know what she believes on some matter? How can we best interpret the expressions used in the quotations above?

3. Explaining the expressions

I'll begin by addressing some potential objections to the view I am sketching here. I am proposing that these expressions about self-knowledge (and its lack) are about *transparent* self-knowledge. I am proposing that we explain the connection between constancy in one's attitudes and self-knowledge by understanding the way that

---

affairs seamlessly into close third-person narration that describes that state of affairs. See Wood (2008).

[217] Montaigne, Michel (2003).

[218] Montaigne (2003), pp.373-7.

inconstancy frustrates your attempts to gain diachronic transparent self-knowledge. You might, at this point, resist the direction of this discussion and try to explain the expressions I have excerpted above in other ways. Let's consider a couple alternative interpretations of the expressions listed above, then, in order to rule them out.

You might, at the outset, suggest that the expressions about self-knowledge are really drawing a connection between knowledge *of* your own inconstancy and your inconstancy. If Lear knew his inconstancy about Cordelia and Kent, or if Mrs. Arbuthnot (as Gerald sees her), knew her inconstancy on the matter of whether Gerald should take the appointment, then they would be motivated to become more constant. To say of one of them that he or she *doesn't know his/her own mind* is just to point out the circumstance that allows their inconstancy to thrive: their lack of knowledge of the inconstancy itself.

This proposal relies on the claim that knowledge of your own inconstancy would motivate you to become more constant, but that claim is not very plausible. As noted above, not all inconstancy is bad. You can change your mind in response to evolving evidence or shifting reasons to act. In those cases, changing your beliefs or your intentions is sometimes actually obligatory.

You are especially likely in the first-personal case to think of your own inconstancy, when you recognize it, as motivated by a change in reasons. Your being the one who actually changed your attitudes in the way you did may even make it the case that you are *less* likely to see the relevant changes as problematic in any way. Even if you should not have changed your mind in response to shifting reasons or evidence, the first-order errors you make in adjusting your intentions to changes in circumstances may be echoed in second-order approbation of those errors.

It is worth noting that Lear and Mrs. Arbuthnot are very much aware of their own inconstancy (or, in Mrs. Arbuthnot's case, seeming inconstancy) in the relevant cases. Lear is taken aback by what he sees as an abrupt shift in his circumstances: betrayal on Cordelia's part and insubordinacy on Kent's part. He remarks that previously he "loved [Cordelia] most, and thought to set [his] rest / On her kind nursery." He recognizes explicitly that Cordelia is "new adopted to [his] hate." And Mrs. Arbuthnot herself cannot have forgotten the previous position she was forced into, when Gerald urged her in public to make any proper objections to his plan, and she found herself unable to speak freely. The inconstancy itself is not in any way hidden from Lear and Mrs. Arbuthnot, and it is not lack of knowledge of such inconstancy that Goneril, Regan, and Gerald bemoan. It is rather the inconstancy itself.

Lear can know what he believes and intends at any one moment, and know precisely how his beliefs and intentions have changed over time, and yet still be said to lack self-knowledge in this context. The same is true of Mrs. Arbuthnot, and for anyone else besides. If you change your mind quite rapidly, even if you are perfectly aware of your changes of mind and the states of mind between which you vacillate, you can still be truly said to lack self-knowledge—in particular, to lack knowledge of what you believe or what you intend to do. This is important to note when we are trying to interpret the expressions discussed above, although it also makes them even more puzzling.

To solve the puzzle at this point, we need to recognize the distinction between *any* self-knowledge of one's attitudes and *transparent* self-knowledge of one's attitudes. At any one point in time in your vacillation, you might know what you believe or intend at that moment, transparently. But if you know what you *did* believe or intend in the past, it

is not (solely) by use of the transparency method; you might need to *remember* what you thought in the past, for example.

We can better interpret the relevant expressions in terms of transparent self-knowledge. What Lear and (as Gerald sees her) Mrs. Arbuthnot lack is *transparent* self-knowledge of *all* that they believe and intend over time. That is, they lack diachronic transparent self-knowledge.[219]

This might seem to be a strange or stretched interpretation. Transparency of belief and intention are not explicitly at issue in the relevant contexts, and Goneril, Regan, and Gerald do not press their criticisms explicitly in terms of *transparent* self-knowledge. However, this worry is not as pressing when we recall a key fact that we already needed to use to explain first-person authority about belief and intention: we presume of one another that we use the transparency method by default to self-attribute beliefs and intentions. This *default* presumption is built into the charges of lack of self-knowledge made by Regan and Gerald in their respective contexts.

The proposal, more rigorously put, is this: to say of someone who is inconstant in her beliefs or intentions that she *doesn't know herself, doesn't know her mind*, or *doesn't know what she believes/intends* is just to say that any momentary use of the transparency method would not yield self-knowledge of all her diachronically shifting beliefs and intentions. There is a hidden restriction, in these expressions, to transparent self-knowledge. The fact that this restriction is implicit, rather than explicit, is explained by our tendency to presume of one another that we use transparent self-attributions by default to self-attribute beliefs and intentions. The claim that we make this restriction in these cases is not *ad hoc*. It must already be the case to explain first-person authority at all that we presume default *transparent* self-attribution of belief and intention.

Once we accommodate this hidden restriction, the expressions discussed above start to make more sense. We can also make sense of distinctions between agents who are said to know themselves well and those who are said to lack self-knowledge. Whereas Lear and Mrs. Arbuthnot (as Gerald sees her) lack diachronic transparent self-knowledge, Edmund Bertram has particularly secure diachronic transparent self-knowledge. The transparent self-attributions of intention that he could make at any one time would capture his steadfast—that is, *constant*—intentions that last over time as well. He is, in that respect, much like Cato the younger, as discussed by Montaigne: "strike one of his keys and you have struck them all; there is in him a harmony of sounds in perfect concord such as no one can deny."[220]

Resolving this interpretive puzzle does not complete the task at hand, though. We can now see why those who are inconstant are said to lack self-knowledge: they lack diachronic *transparent* self-knowledge, when diachronic self-knowledge is at issue. But we have not yet seen why constancy itself should be valuable. I'll turn to this issue next.

---

[219] It is clear that diachronic self-knowledge is at issue in these contexts, because Lear and Mrs. Arbuthnot are said not to know their *minds* or *selves*, which are taken to endure through time. Note, though, that Lear could equally be said not to know what he believes, and Mrs. Arbuthnot could be said not to know what she intends.

[220] Montaigne (2003), p.375.

4. The value of constancy

This chapter explains the value of certain kinds of transparent self-knowledge. Above, I argued that complete transparent self-knowledge is valuable insofar as having *that* requires having no hidden attitudes. Now we are considering whether *diachronic* transparent self-knowledge is valuable. We have seen that diachronic transparent self-knowledge requires some measure of constancy in those attitudes you can transparently self-attribute—i.e. belief and intention. But we have not yet seen why it is valuable to have constancy in belief or constancy in your intentions. I'll now ask why constancy of this kind is valuable, both to complete our inquiry into the value of transparent self-knowledge, and to finish our interpretation of the forms of praise and criticism I addressed in the last two sections of this paper.

There is no straightforward way to answer this question, because constancy is not required to have beliefs and intentions that fulfill other normative requirements. You can have warranted beliefs, and reasonable intentions, even though they are not constant over time. In fact, in circumstances in which your reasons to believe and your reasons to act are shifting over time, you would do better to change your mind about what is true and what is best to do. Sometimes the demands of reason recommend inconstancy.

In his extended discussion of constancy as a virtue, Montaigne proposed a closely related view about its value. He approvingly quoted Seneca ("an Ancient") on this point:

> "Wisdom," said an Ancient, "is always to want the same thing, always not to want the same thing." I would not condescend to add, he said, "provided that your willing be right. For if it is not right, it is impossible for it to remain ever one and the same."
>
> I was once taught indeed that vice is no more than a defect and irregularity of moderation, and that consequently it is impossible to tie it to constancy. There is a saying attributed to Demosthenes: the beginning of all virtue is reflection and deliberation: its end and perfection, constancy.[221]

The idea here is that it is only possible to gain great constancy in your attitudes when you have recognized what is right, what is good, and what is true. I don't think we need much argument to dismiss this rosy picture of the world. Demagogues, tyrants, and those incapacitated by delusion can be just as constant as anyone would like to be.[222] It is simply not the case that constancy in belief implies the truth of those beliefs. Nor is it the case that constancy in intention implies the rightness of those intentions.[223]

We have seen one way in which having constant beliefs and intentions gives you an epistemic advantage. Transparent self-knowledge of what you believe or intend at any given moment is simple to gain. When your beliefs and intentions are constant, you need

---

[221] Montaigne (2003), p.374. The quotation, ed. Screech notes, is from Seneca's *Epistles* XX.5.

[222] We might, in those circumstances, use a negatively valenced term like "inflexible" or "stubborn" (compare my use above) to indicate this quality in a contemptible context. Still, being stubborn or inflexible just *is* a way of being constant.

[223] It may be true that having only true beliefs implies having constant beliefs, insofar as the truths of various matters don't change. This proposal does not help us much in this context, as this is not a proposal about the value of *constancy* at all.

do no more than self-attribute beliefs and intentions all at one moment in order to self-attribute beliefs and intentions over time as well. It is more difficult, and sometimes much more difficult, to gain diachronic self-knowledge of your beliefs and intentions when they are inconstant. You must track and remember how your attitudes change in order to have diachronic self-knowledge when you are inconstant. Diachronic self-knowledge is simpler, then, for those whose attitudes are constant.

But this cannot be the whole story about the value of constancy in your attitudes. When Goneril and Regan criticized their father for his inconstancy, they did not seem primarily concerned with the ease with which he might gain self-knowledge. Nor did Gerald Arbuthnot have any particular concerns about his mother's epistemic efforts when he complained about the change in her recommendations about what he should do. If the criticisms about self-knowledge leveled in these contexts have anything to do with the value of constancy (or the disvalue of inconstancy), we should have more to say about why it is good to retain the same beliefs and intentions over time.

I have not here surveyed every proposal about the value of constancy in our attitudes. To avoid getting lost in a mass of proposals, I will move to make my positive proposals now, instead of summarizing other views as well. I will make two such positive proposals about the value of constancy. The first is that constancy of belief and constancy of intention can be valuable for local reasons in individual contexts. The second is that some constancy in your attitudes is necessary to be a diachronically unified self.

The local value that constancy has changes from context to context to some extent. In Lear's context, constancy in his attitudes is valuable to those who want to serve his wishes and avoid offending him. In Gerald's context, constancy in his mother's attitudes is required for him to settle on a stable plan that would please her. The constancy in Edmund's intentions makes him a predictable and dependable friend and fiancé.

It is difficult to deny that constancy in these contexts has simple utility of this kind. Constancy is useful to those who want to plan around you—whether they mean to please you or to anger you. Constancy supports coordination in various different contexts.[224] But this does not seem to be the end of the story either. It seems that constancy should be valuable in some way for the person who is constant, and inconstancy should be problematic for the person who is inconstant. The proposal about coordination is largely a proposal about how constancy is valuable from the second- or third-personal perspective.

The second positive proposal I will make is a proposal about why constancy is valuable to the person who has it. In order to be a diachronically unified self, it seems that you must have a certain measure of constancy in your attitudes, including (but not limited to) belief and intention.

To clarify what I mean by a diachronically unified self, and to demonstrate that we really do value having such diachronically unified selves, I'll turn to consider the poem "Self-Knowledge" written in 1832 by Samuel Taylor Coleridge.[225] Here, in opposition to the famous ancient exhortation to *know thyself*, inscribed on the temple at Delphi, Coleridge cautions you against futile attempts to know yourself:

---

[224] That is not to imply that constancy is always best for coordination. Sometimes when you are coordinating with a partner—say, on a doubles team in tennis—you need to depend on their beliefs and intentions changing in step with yours.

[225] Coleridge (1832).

> νῶθι σεαυτόν![226]—and is this the prime
> And heaven-sprung adage of the olden time!—
> Say, canst thou make thyself?—Learn first that trade;—
> Haply thou mayst know what thyself had made.
> What hast thou, Man, that thou dar'st call thine own?—
> What is there in thee, Man, that can be known?—
> Dark fluxion, all unfixable by thought,
> A phantom dim of past and future wrought,
> Vain sister of the worm,—life, death, soul, clod—
> Ignore thyself, and strive to know thy God!

The idea is that there is a sense in which you can *fail* to meet a basic condition on being knowable, even to yourself. If your mind is characterized by "Dark fluxion, all unfixable by thought," the most genuine question to ask is not *how* to know yourself, but rather "What is there in thee … that can be known?" In a condition of constant "fluxion"—what Montaigne called "vacillation"—the task of self-knowledge is not just difficult, but fundamentally misguided. There is a real puzzle about whether there is anything to "call thine own" in a mind that keeps changing. Any conception of yourself as a lasting thing seems nothing but an abstraction, a "phantom dim of past and future wrought."

Montaigne, in his discussion of constancy, had often expressed the same worry:

> given the natural inconstancy of our behavior and our opinions it has often occurred to me that even sound authors are wrong in stubbornly trying to weave us into one invariable and solid fabric.[227]

It is a classic move on Montaigne's part to use excerpts from the ancients to express what most agree is a distinctively modern thought about the ways in which the most ordinary of selves might splinter.[228] Here, again, Montaigne puts Seneca to work for his point:

> there is as much difference between us and ourselves as there is between us and other people. '*Magnam rem puta unum hominem agere*' [Let me convince you that it is a hard task to be always the same man.][229]

Though both Montaigne and Coleridge seem to agree that you might *fail* to have a self in

---

[226] "νῶθι σεαυτόν" (*gnothi seauton*), or "know thyself", was inscribed at the temple at Delphi. The poem begins with the epigraph "E coelo descendit γνῶθι σεαυτόν.—JUVENAL, xi. 27." The title "Self-Knowledge" first appeared with the poem in 1893.

[227] Montaigne (2003), p.373.

[228] For a useful discussion of what made Montaigne the first modern philosopher of the self, see Taylor (1989), Part II, Chapter 10, "Exploring 'L'humaine condition.'" Taylor writes: "there is some evidence that when [Montaigne] embarked on his reflections, he shared the traditional view that these should serve to recover contact with the permanent, stable, unchanging core of being in each of us. This is the virtually unanimous direction of ancient thought: beneath the changing and shifting desires in the unwise soul, and over against the fluctuating fortunes of the external world, our true nature, reason, provides a foundation, unwavering and unconstant" (p.178).

[229] Montaigne (2003), p.380. The Seneca quotation is from *Epistulae morales* CXX.22.

some substantive, interesting way, they disagree vastly on the practical upshot of this claim. Montaigne counsels us to give up the search for self-knowledge: he "would that fewer people would concern themselves" with the "chancy undertaking" of searching for principles that underlie their thoughts and actions as a whole.[230]

Coleridge, on the other hand, is more equivocal—in a way that exhibits, in the very form of his short verse, the very instability he means to describe. He ends by counseling the reader to *ignore* herself, at least as a potential epistemic object. But he has already suggested that there is something else you might do, instead of chasing down knowledge of something that doesn't even exist for you to know. "Say, canst thou make thyself?" he asks. "Learn first that trade;— / Haply thou mayst know what thyself had made." Self-making would involve something like making up your mind, or shaping your character, into something that would be knowable in a way you are not now knowable.

It is this kind of self-making that is required to go from being someone without diachronic transparent self-knowledge to being someone with diachronic transparent self-knowledge. If you are not unified over time, there is in some meaningful sense no lasting self there for you to get to know. There is a sense in which the inconstant among us *lack* a kind of selfhood enjoyed by those who are constant.

We need not take literally Montaigne's point that there is "as much difference between us and ourselves as there is between us and other people" in order to make sense of some philosophically respectable notion of a self that draws some such empirical distinction. Consider some of the ways we use this kind of talk to capture changeability in our friends and colleagues. You might say "she's one person around me, and an entirely different person around her parents." Or you might claim, of a friend who has decided to run for office and has abandoned her previous values in the process: "that's not the person I know." Or, cringing at a paper you wrote in the past, you might reassure yourself: "I was a different person back then."[231]

We could easily take these claims to be metaphorical, or rough, in a way that rejects the imputation that there *really is* no one unified self in one human body. I don't think that is the right reaction to these formulations, though. The right reaction is to admit a more nuanced understanding of the type of self, or person, we mean to describe when we use these kinds of expressions. Each of us whose attitudes have sufficient constancy have a self that is individuated by her beliefs, desires, intentions, hopes, values, and other attitudes that guide thought and behavior.[232] Those of us who are inconstant in relevant ways fail to have unified selves in this sense. To fail to *know yourself* on this understanding of a self is to fail to *be* one self that is knowable by transparent means.

It might seem that the connection between constancy and being a unified self has just pushed back the question about value. Why should we value being unified selves? Even if it does appear—in Coleridge's exhortations, and in Montaigne's aspersions—that we do value having diachronically unified selves, are we right to do so?

This is a deeper question than I can properly address in the remainder of this chapter. There are, however, a few things to be said in favor of having a unified self. Having a unified self might be a precondition on evaluation of important kinds. In particular, it seems that there is not much sense to be made of the property of authenticity for someone

[230] Montaigne (2003), p.380.
[231] Compare: "I'm of two minds about this matter," or "I have half a mind to do it."
[232] Compare Korsgaard (1996) on practical identity.

who is so inconstant as to lack a unified self. In order to qualify as authentic, *or* inauthentic, you must have a unified self against which to measure your actions and thoughts. Those actions and thoughts cannot even be evaluated by the standard of authenticity if there is no unified self to which you be authentic or fail to be authentic.

You cannot, in serious inconstancy, act in a way that expresses your one true point of view. If you are inconstant, your actions are constrained to flow from temporary or infirm motives, rather than resonating with a unified self. That is an unhappy fate. Even if you cannot be *inauthentic* in this situation, neither can you ever hope to be authentic.

Constancy is valuable, then, not just to others who want to plan around you. It is also valuable to *you* insofar as you value having a diachronically unified self, and thus being evaluable by standards of authenticity.

Conclusion

What does all this mean for the value of transparent self-knowledge? I have not argued in this chapter that transparent self-knowledge in general is valuable. I have argued that *complete* and *diachronic* transparent self-knowledge is valuable insofar as having it involves having other goods. To have complete transparent self-knowledge is to lack hidden attitudes that resist conscious consideration and revision. To have diachronic transparent self-knowledge is also to have constancy in your attitudes; to have sufficient constancy in your attitudes is to have a diachronically unified self. We could say, then, that transparent self-knowledge of the complete and diachronic kind has **emblematic value**: it is valuable insofar as having it involves having (only) transparent attitudes and having a diachronically unified self.

It may still remain puzzling why we value constancy and unified selves. I have tried to allay this puzzlement with reference to literary quotations that demonstrate at the very least that we *do* value constancy and unification over time. I have also gestured at the significance of diachronic unification by connecting it with authenticity: you fail to meet a precondition on evaluation for authenticity when you lack a diachronically unified self.

It is worth emphasizing an important methodological point in closing. The interpretive exercises of this chapter, which aimed to interpret various literary expressions in terms of *transparent* self-knowledge, relied on a fact that has also served as a touchstone in past chapters of this dissertation. That key fact is that we presume of one another that we use transparent self-attribution of belief and intention by default. Without this presumption, we could not make sense of the hidden restriction to transparent self-knowledge in phrases like "she doesn't know what she believes" and "she doesn't know what she intends to do." You can be truly said to lack *this* kind of self-knowledge—that is, diachronic transparent self-knowledge—even when you have *some* kind of knowledge of all the facts about your attitudes over time.

That we can speak of someone lacking self-knowledge although she knows all the facts there are to know about herself is a remarkable consequence of this chapter's discussion. It also emphasizes the fact that the task of gaining such self-knowledge is not the distinctively epistemic task of figuring out the way that you already are. It is, rather, the task of shaping yourself into the kind of self that is knowable in this special, transparent way: a diachronically unified self with no hidden beliefs and intentions.

Conclusion. Mental Action and Self-Knowledge

In this dissertation I have explained how you know what you believe and how you know what you intend to do. I have argued that you can self-attribute both beliefs and intentions transparently. Any transparent self-attribution of a belief or intention constitutes authoritative knowledge of your belief or intention—at least at the moment you make it. This view of transparent self-knowledge also provides general necessary and sufficient conditions on any mental act or state *M*'s being transparently self-attributable.

This view of transparent self-knowledge also provided a solution to Moore's paradox: the paradox of explaining why certain contents involving belief attributions are absurd to judge or to assert, although these contents are satisfiable. I explained that using the same facts that contribute to an explanation of transparent self-knowledge of belief.

However, we saw that transparent self-knowledge can come apart from Moorean absurdities. There are contents involving *intention* attributions that are in some way absurd to assert or to judge even though they are satisfiable. But these contents are not properly *Moorean* absurdities: the explanation of their absurdity does not proceed along the same lines as the explanation in the case of belief.

Analysis of transparent self-knowledge reveals that it is epistemically groundless. However, epistemic groundlessness is not special to the first-person perspective. Interpretation of others' behavior also requires you to make implicit belief attributions to other people that are themselves epistemically groundless.

The immediacy and simplicity of transparent self-knowledge does not imply that it is not valuable in any way. Complete and diachronic transparent self-knowledge has emblematic value. To have that is also to lack hidden attitudes, and to have a diachronically unified self.

Those are the main conclusions of this dissertation. In the Introduction, I stated that I wanted to explain all of the following:

- why we usually know what we believe and what we intend
- how we can fail to know our beliefs and intentions
- how first-personal and third-personal methods of attributing these states compare
- why self-knowledge of beliefs and intentions is important to us personally

In drawing the conclusions just described, I have answered these questions.

In order to answer these questions, I had to argue for some further important conclusions, whose importance extends beyond the scope of this dissertation. I have argued that practical knowledge depends on control in the way that empirical knowledge depends on justification. I have argued that you have strong control over the attitudinal aspect of your thought. I have argued that embedding some mental actions in overarching mental tasks allows one and the same mental action to have distinct contents under distinct intentional descriptions that apply to it. I have argued that the lack of truly Moorean absurdities for intention speaks to the fundamental difference between the attitudinal aspect of belief and intention: intention is not a form of belief. I have also argued that constancy in our attitudes is valuable insofar as it allows you to have a diachronically unified self, which itself is required for authenticity (and inauthenticity).

The most important point in this dissertation is that we cannot understand the

epistemology of self-knowledge without understanding the way that we can perform actions in thought. Our ability to perform intentional mental actions solves the attitude problem for self-knowledge of belief and for self-knowledge of intention. It explains how some thoughts can have content plurality. It helps solve Moore's paradox. It also explains the groundlessness of our first-personal self-attributions of belief and intention—even though this feature is not special to *first*-personal attributions of belief and intention.

This point is not the same as the more general claim that agency is important to self-knowledge. That claim is true, but it does not do nearly enough to specify why agency is important to self-knowledge, and what we would lack without it. Understanding intentional mental action in particular goes much further in this regard.

I gave this dissertation the title "Knowing Yourself is Something You Do." This title is meant to affirm that you really do know your mind in the way that you are ordinarily taken to know your mind—by yourself and others. But it is mainly meant to emphasize, by way of a slight infelicity in grammar, that the knowledge you have of your own mind is in large part the kind of knowledge that an *agent* has of what she is doing. If you could not do things in thought, you would not know your mind, or your *self*, as well as you do.

References

"Shall." Entry in *The Oxford English Dictionary* online.

Albritton, Rogers (1995). "Comments on 'Moore's paradox and self-knowledge." *Philosophical Studies* 77.2/3: 229-239.

Anscombe, G.E.M. (1957). *Intention*. Oxford: Basil Blackwell.

Austen, Jane (1816/2001). *Mansfield Park*. Dover: Mineola, New York.

Austin, J.L. (1962). *How to Do Things With Words*. Oxford: Clarendon Press.

Barnett, D. J. (2015). "Inferential justification and the transparency of belief." *Noûs* 50.1: 1-29.

Bennett, Jonathan (1988). *Events and their Names*. Indianapolis: Hackett.

Bilgrami, Akeel (1998). *Self-Knowledge and Resentment*. Cambridge, MA: Harvard University Press.

Boghossian, Paul (1989). "Content and self-knowledge." *Philosophical Topics* 17.1: 5-26.

Boghossian, Paul (2003). "The normativity of content." *Philosophical Issues* 13: 31-45.

Boghossian, Paul (2014). "What is inference?" *Philosophical Studies* 169: 1-18.

Boghossian, Paul and Timothy Williamson (2003). "Blind reasoning." *Proceedings of the Aristotelian Society* 77: 225-293.

Boyle, Matthew (2009a). "Active belief." *Canadian Journal of Philosophy* 39 sup.1: 119-147.

Boyle, Matthew (2009b). "Two kinds of self-knowledge." *Philosophy and Phenomenological Research* 78: 133-63.

Boyle, Matthew (2011a). "Making up your mind and the activity of reason." *Philosopher's Imprint* 11.17.

Boyle, Matthew (2011b). "Self-knowledge and transparency II: Transparent self-knowledge." *Proceedings of the Aristotelian Society* Supp. Vol. LXXXV: 223-241.

Bratman, Michael (1999). *Intention, Plans, and Practical Reason*. Stanford, CA: Center for the Study of Language and Information.

Brueckner, Anthony (1998). "Moore inferences." *The Philosophical Quarterly* 48.192: 366-369.

Burge, Tyler (1996). "Our entitlement to self-knowledge I." *Proceedings of the Aristotelian Society* 96: 91-116.

Byrne, Alex (2005). "Introspection." *Philosophical Topics* 33.1: 79-104.

Byrne, Alex (2011). "Self-knowledge and transparency I: Transparency, belief, intention." *Proceedings of the Aristotelian Society* Supp. Vol. LXXXV: 201-221.

Byrne, Alex (2012). "Knowing what I want." In Jeeloo Liu and John Perry, eds., *Consciousness and the Self: New Essays,* pp. 165-183. New York: Cambridge University Press.

Cassam, Quassim (2014). *Self-Knowledge for Humans*. Oxford: Oxford University Press.

Chalmers, David (2003). "The content and epistemology of phenomenal belief." In Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Perspectives*. New York: Oxford University Press.

Churchland, Paul M. (1981/2002). "Eliminative materialism and the propositional attitudes." In David Chalmers, ed., *Philosophy of Mind: Classical and Contemporary Readings*. New York: OUP. 568-580.

Conee, E. and R. Feldman (1998). "The generality problem for reliabilism." *Philosophical Studies* 89.1: 1-29.
Coleridge, Samuel Taylor (1832/2009). "Self-knowledge." In *The Complete Poetical Works of Samuel Taylor Coleridge*, ed. Ernest Hartley Coleridge. http://www.gutenberg.org/files/29090/29090-h/29090-h.htm Accessed May 1, 2017.
Crane, Tim (2001). *Elements of Mind: An Introduction to the Philosophy of Mind*. New York: Oxford University Press.
Davidson, Donald (1963/2001). "Actions, reasons, and causes." In Donald Davidson, *Essays on Action and Events* (pp.3-19). Oxford: Clarendon Press.
Davidson, Donald (1967/2001). "The logical form of action sentences." In *Essays on Actions and Events* (Oxford: Clarendon Press), 105-121.
Davidson, Donald (1971/2001). "Agency." In *Essays on Actions and Events* (Oxford: Clarendon Press), 43-62.
Davidson, Donald (1973/2001a). "Freedom to act." In *Essays on Action and Events* (Oxford: Clarendon Press), 63-81.
Davidson, Donald (1973/2001b). "Radical interpretation." In *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press), 125-140.
Davidson, Donald (1974a/2001). "Belief and the basis of meaning." In *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press), 141-154.
Davidson, Donald (1974b/2001). "On the very idea of a conceptual scheme." In Donald Davidson, *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
Davidson, Donald (1974c/2001). "Replies to David Lewis and W.V. Quine." In Donald Davidson, *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press. 280-285.
Davidson, Donald (1978/2001). "Intending." In *Essays on Actions and Events* (Oxford: Clarendon Press), 83-102.
Davidson, Donald (1984/2001). "First person authority." *Dialectica* 38.2/3: 101-111.
Davidson, Donald (1987/2001). "Knowing one's own mind." In Donald Davidson, *Subjective, Intersubjective, Objective* (pp. 15-38). Oxford: Clarendon Press.
Davidson, Donald (1991/2001). "Three varieties of knowledge." In Donald Davidson, *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press.
Davies, Martin (1982). "Idiom and metaphor." *Proceedings of the Aristotelian Society* 83: 67-85.
De Almeida, C. (2001). "What Moore's paradox is about." *Philosophy and Phenomenological Research* 62: 33-58.
Dennett, Daniel (1981/2002). "True believers: The intentional strategy and why it works." In David Chalmers, ed., *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press. 556-568.
Dennett, Daniel (1991). "Real patterns." *Journal of Philosophy* 81.1: 27-51.
DeRose, Keith (1999). "Contextualism: An explanation and defense." In John Greco and Ernest Sosa, eds., *The Blackwell Guide to Epistemology* (Hoboken, NJ: Blackwell Publishers), 187-205.
Dorsch, Fabian (2009). "Judging and the scope of mental agency." In Lucy O'Brien and Matthew Soteriou, eds., *Mental Actions* (New York: Oxford University Press), 38-71.
Dretske, Fred (2003a). "How do you know you are not a zombie?" In *Privileged Access and First-Person Authority*, ed. B. Gertler. Aldershot: Aldershot Publishing.
Dretske, Fred (2003b). "Knowing what you think vs. knowing that you think." In *The*

Externalist Challenge: New Studies on Cognition and Intentionality, ed. R. Schantz. Berlin: Walter de Gruyter.
Dretske, Fred (2012a). "Awareness and authority: Skeptical doubts about self-knowledge." In Declan Smithies and Daniel Stoljar, eds., *Introspection and Consciousness*. New York: Oxford University Press.
Dretske, Fred (2012b). "I think I think, therefore I am – I think: skeptical doubts about self- knowledge." In Jeeloo Liu and John Perry, eds., *Consciousness and the Self: New Essays*. New York: Cambridge University Press.
Edgley, Roy (1969). *Reason in Theory and Practice*. London: Hutchinson.
Evans, Gareth (1982). *The Varieties of Reference*, ed. John McDowell. New York: Oxford University Press.
Ford, Anton, Jennifer Hornsby, and Frederick Stoutland, eds. (2014). *Essays on Anscombe's Intention*. Cambridge, MA: Harvard University Press.
Frege, Gottlob (1956). "The thought: A logical inquiry." Trans. A.M. and Marcelle Quinton. *Mind* 65.259: 289-311.
Frege, Gottlob. (1979). "Logic." In *Posthumous Writings*. Chicago: University of Chicago Press.
Fricker, Elizabeth (1998). "Self-knowledge: Special access versus artefact of grammar—a dichotomy rejected." In Crispin Wright, Barry C. Smith, and Cynthia MacDonald, eds., *Knowing Our Own Minds*. Oxford: Clarendon Press. 155-206.
Gallois, André (1996). *The World Without, the Mind Within*. New York: Cambridge University Press.
Geeraert, Kristina, John Newman, and R. Harald Baayen (2017). "Idiom variation: Experimental data and a blueprint of a computational model." *Topics in Cognitive Science*
Gertler, Brie (2011). *Self-Knowledge*. New York: Routledge.
Gertler, Brie (2012). "Renewed acquaintance." In Declan Smithies and Daniel Stoljar, eds., *Introspection and Consciousness* (New York: Oxford University Press), 94-127.
Gertler, Brie (forthcoming). "Self-knowledge and rational agency: A defense of empiricism." *Philosophy and Phenomenological Research.*
Gettier, Edmund (1963). "Is justified true belief knowledge?" *Analysis* 23(6), 121-123.
Gibbard, Alan (2003). "Thoughts and Cais." *Philosophical Issues* 13: 83-98.
Gibbard, Alan (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.
Goldman, Alvin (1976). "Discrimination and perceptual knowledge." *Journal of Philosophy* 73: 771-91.
Goldman, Alvin (1979). "What is justified belief?" In George Pappas, ed., *Justification and Knowledge* (Dordrecht: D. Reidel), 1-25.
Goldstein, Laurence (1993). "Inescapable surprises and acquirable intentions." *Analysis* 53.2: 93-99.
Gopnik, Alison (1993). "How we know our minds: the illusion of first-person intentionality." *Brain and Beahvioral Sciences* 16: 1-14.
Gordon, Robert (1986). "Folk psychology as simulation." *Mind & Language* 1.2: 158-171.
Green, Mitchell (2007). "Moorean absurdity and showing what's within." In Mitchell Green and John N. Williams, eds., *Moore's Paradox.* Oxford: Clarendon Press.
Green, Mitchell and John N. Williams, eds. (2007). *Moore's Paradox: New Essays on*

Belief, Rationality, and the First Person. New York: Oxford University Press.
Hall, Ned (2004). "Two concepts of causation." In John Collins, Ned Hall and Laurie Paul, eds., *Causation and Counterfactuals* (Cambridge, MA: MIT Press), 225-276.
Hampshire, Stuart (1959). *Thought and Action*. London: Chatto and Windus.
Heal, Jane (1994). "Moore's paradox: A Wittgensteinian approach." *Mind* 103.409: 5-24.
Heal, Jane (2002). "The presidential address: On first-person authority." *Proceedings of the Aristotelian Society*, New Series, 102: 1-19.
Heil, John (1998). "Privileged access." *Mind* 97: 238–251.
Hintikka, Jaakko (1962). *Knowledge and Belief*. Ithaca, NY: Cornell University Press.
Horgan, Terry and Uriah Kriegel (2007). "Phenomenal epistemology: What is consciousness that we may know it so well?" *Philosophical Issues* 17: 123-144.
Howell, Robert J. (2008). Review of Lucy O'Brien's *Self-Knowing Agents*. *Notre Dame Philosophical Reviews* 2008.03.21.
Hume, David (1978). *A Treatise of Human Nature*. 2nd ed. Ed. L.A. Selby-Bigge and P.H. Nidditch. Oxford: Clarendon Press, 1978.
Hyslop, Alec (2014). "Other inds." *Stanford Encyclopedia of Philosophy*. Online: http://plato.stanford.edu/entries/other-minds/
Kant, Immanuel (1998). *Critique of Pure Reason*. Trans. Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.
Kavka, Gregory (1983). "The toxin puzzle." *Analysis* 43.1: 33-36.
Kitcher, Patricia (2011). *Kant's Thinker*. New York: Oxford University Press.
Korsgaard, Christine (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.
Kriegel, Uriah (2004). "Moore's paradox and the structure of conscious belief." *Erkenntnis* 61: 99-121.
Kripke, Saul (2011). "Nozick on knowledge." In *Philosophical Troubles: Collected Papers, Volume I* (New York: Oxford University Press), 162-225.
Kripke, Saul (1982). *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.
Larkin, William S. (1999). "Shoemaker on Moore's paradox and self-knowledge." *Philosophical Studies* 96.3: 239-252.
Lear, Jonathan (2004). "Avowal and unfreedom." *Philosophy and Phenomenological Research* 69.2: 448-454.
Lewis, David (1974). "Radical interpretation." *Synthese* 27: 331-344.
Lewis, David (1982). "Logic for equivocators." *Noûs* 16.3: 431-441.
Lewis, David (1986). *On the Plurality of Worlds*. Malden, MA: Blackwell Publishing.
Longuenesse, Béatrice (2001). *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*. Princeton: Princeton University Press.
MacFarlane, John (2011). "What is assertion?" In Jessica Brown and Herman Cappelen, eds., *Assertion*. New York: Oxford University Press.
Malcolm, Norman (1958). "Knowledge of other minds." *JPhil* 55: 35-52.
Martinich, A. (1980). "Conversational maxims and philosophical problems." *Philosophical Quarterly* 30: 215-28.
McGinn, Colin (1982). *The Character of Mind: An Introduction to the Philosophy of Mind*. New York: OUP.

Montaigne, Michel de (2003). "On the inconstancy of our actions." In *The Complete Essays*. Trans. M.A. Screech (Penguin Books: London), 373-380.
Moore, Michael S. (2010). *Act and Crime: The Philosophy of Action and its Implications for Criminal Law*. New York: Oxford University Press.
Moore, G.E. (1993) "Moore's paradox." In Thomas Baldwin, ed., *G.E. Moore: Selected Writings* (New York: Routledge), 207-212.
Moore, G. E. (1993). *Selected Writings*, ed. T. Baldwin. London: Routledge.
Moran, Richard (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
Moran, Richard (2003). "Responses to O'Brien and Shoemaker." *European Journal of Philosophy* 11.3: 402-419.
Moran, Richard (2004a). "Replies to Heal, Reginster, Wilson, and Lear." *Philosophy and Phenomenological Research* 69.2: 455-472.
Moran, Richard (2004b). "Anscombe on 'practical knowledge.'" In J. Hyman and H. Steward, eds., *Agency and Action* (New York: Cambridge University Press), 43-68.
Nozick, Robert (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
O'Brien, Lucy (2003). "Moran on agency and self-knowledge." *European Journal of Philosophy* 11.3: 375-390.
O'Brien, Lucy (2005). "Self-knowledge, agency, and force." *Philosophy and Phenomenological Research* 71.3: 580-601.
O'Brien, Lucy (2007). *Self-Knowing Agents*. New York: Oxford University Press.
O'Brien, Lucy and Matthew Soteriou, eds. (2009). *Mental Actions*. New York: Oxford University Press.
Olson, Christopher (1969). "Knowledge of one's own intentional actions." *The Philosophical Quarterly* 19.77: 324-336.
Paul, Sarah K. (2009a). "How we know what we're doing." *Philosophers' Imprint* 9.11.
Paul, Sarah K. (2009b). "Intention, belief, and wishful thinking: Setiya on 'practical knowledge.'" *Ethics* 119.3: 546-557.
Paul, Sarah K. (2012). "How we know what we intend." *Philosophical Studies* 161, 327-346.
Peacocke, Antonia (2017). "Embedded mental action in self-attribution of belief." *Philosophical Studies* 174.2: 353-377.
Peacocke, Christopher (1996). "Our entitlement to self-knowledge II: Entitlement, self-knowledge, and conceptual redeployment." *Proceedings of the Aristotelian Society* 96: 117-158.
Peacocke, Christopher (1998). "Conscious attitudes, attention, and self-knowledge." In Crispin Wright, Barry C. Smith, and Cynthia MacDonald, eds., *Knowing Our Own Minds* (pp.63-121). Oxford: Clarendon Press, 1998.
Peacocke, Christopher (2008). "Mental action." In Christopher Peacocke, *Truly Understood* (pp.245-285). New York: Oxford University Press.
Pitt, David (2004). "The phenomenology of cognition or *what is it like to think that P*?" *Philosophy and Phenomenological Research* LXIX.1: 1-36.
Pryor, James (1999). "Immunity to error through misidentification." *Philosophical Topics* 26(1), 271-304.
Pryor, James (2003). "There is immediate justification." In Matthias Steup and Ernest

Sosa, eds., *Contemporary Debates in Epistemology*. Malden, MA: Blackwell, 2005.
Quine, W.V.O. (1960). *Word and Object.* Cambridge, MA: MIT Press.
Railton, Peter (2006). "Moral factualism." In *Contemporary Debates in Moral Theory*, ed. James Dreier. (Malden, MA: Blackwell Publishing), p.201-219.
Reginster, Bernard (2004). "Self-knowledge, responsibility, and the third person." *Philosophy and Phenomenological Research* 69.2: 433-439.
Russell, Bertrand (1910/11). "Knowledge by acquaintance and knowledge by description." *Proceedings of the Aristotelian Society* 11.5: 108-128.
Russell, Bertrand (1912). *The Problems of Philosophy*. Online: Project Gutenberg.
Russell, Bertrand (1948). "Analogy." In *Human Knowledge: Its Scope and Limits*. Crows Nest: George Allen and Unwin, 482-486.
Ryle, Gilbert (1949). *The Concept of Mind*. London: Barnes & Noble.
Ryle, Gilbert (1971a). "A puzzling element in the notion of thinking." In Gilbert Ryle, *Collected Papers, Volume II: Collected Essays 1929-1968* (pp.391-406). London: Hutchinson.
Ryle, Gilbert (1971b). "Thinking and reflecting." In Gilbert Ryle, *Collected Papers, Volume II: Collected Essays 1929-1968* (pp.465-479). London: Hutchinson.
Ryle, Gilbert (1971c). "The thinking of thoughts: What is 'Le Penseur' doing?" In Gilbert Ryle, *Collected Papers, Volume II: Collected Essays 1929-1968* (pp.480-496). London: Hutchinson.
Schwenkler, J. (2012). "Non-observational knowledge of action." *Philosophy Compass* 7.10: 731-740.
Schwenkler, J. (2015). "Understanding practical knowledge." *Philosopher's Imprint* 15.
Schwitzgebel, Eric (2002). "A phenomenal, dispositional account of belief." *Noûs* 36(2), 249-275.
Schwitzgebel, Eric (2012). "Self-ignorance." In JeeLoo Liu and John Perry, eds., *Consciousness and the Self: New Essays*. New York: Cambridge University Press.
Schwitzgebel, Eric (2015). "Belief." *Stanford Encyclopedia of Philosophy*. Online: http://plato.stanford.edu/entries/belief/
Searle, John and Daniel Vanderveken (1985). *Foundations of Illocutionary Logic*. New York: Cambridge University Press.
Searle, John (1989). "How performatives work." *Linguistics and Philosophy* 12: 535-558.
Setiya, Kieran (2008). "Practical knowledge." *Ethics* 118.3: 388-409.
Setiya, Kieran (2009). "Practical knowledge revisited." *Ethics* 120.1: 128-137.
Setiya, Kieran (2011). "Knowledge of intention." In Anton Ford, Jennifer Hornsby, and Frederick Stoutland, eds., *Essays on Anscombe's Intention* (pp.170-197). Cambridge, MA: Harvard University Press.
Shah, Nishi (2003). "How truth governs belief." *The Philosophical Review* 112.4: 447-482.
Shah, Nishi (2008). "How action governs intention." *Philosopher's Imprint* 8.5: 1-19.
Shah, Nishi and J. David Velleman (2005). "Doxastic deliberation." *The Philosophical Review* 114(4), 497-534.
Shakespeare, William (2005). *The Tragedy of King Lear*: *The Folio Text*. In Stanley Wells and Gary Taylor, eds., *The Oxford Shakespeare: The Complete Works*, 2nd ed (Oxford: Clarendon Press), 1153-1184.
Shepherd, Joshua (2014). "The contours of control." *Philosophical Studies* 170: 395-411.

Shoemaker, Sydney (1968). "Self-reference and self-awareness." *Journal of Philosophy* 65(19), 555-567.
Shoemaker, Sydney (1988). "On knowing one's own mind." *Philosophical Perspectives* 2, 183-209.
Shoemaker, Sydney (1995). "Moore's paradox and self-knowledge." *Philosophical Studies* 77(2/3), 211–228.
Shoemaker, Sydney. (1996). *The First-Person Perspective and Other Essays*. New York: Cambridge University Press.
Shoemaker, Sydney (2003). "Moran on self-knowledge." *European Journal of Philosophy* 11.3: 391-401.
Shoemaker, Sydney (2009). "Self-intimation and second-order belief." *Erkenntnis* 71: 35-51.
Sifferlin, Alexandra (2012). "Top 10 Famous Love Letters." *TIME*. Accessible online at http://newsfeed.time.com/2012/02/14/top-10-famous-love-letters/
Silins, Nicholas (2012). "Judgment as a guide to belief." In Declan Smithies and Daniel Stoljar, eds., *Introspection and Consciousness* (pp.295-327). New York: Oxford University Press.
Sorensen, Roy A. (1988). *Blindspots*. Oxford: Clarendon Press.
Sosa, Ernest (1999). "How to defeat opposition to Moore." *Noûs* 33.13: 141-153.
Soteriou, Matthew (2013). *The Mind's Construction: The Ontology of Mind and Mental Action*. Oxford: Oxford University Press.
Stalnaker, Robert (1984). *Inquiry*. Cambridge, MA: MIT Press.
Stroud, Barry (1984). *The Significance of Philosophical Scepticism*. Oxford: OUP.
Taylor, Charles (1989). *Sources of the Self: The Making of Modern Identity*. Cambridge, MA: Harvard University Press.
Velleman, J. David (1989). *Practical Reflection*. Princeton, NJ: Princeton University Press.
Velleman, J. David (2007). "What good is a will?" In A. Leist, ed., *Action in Context* (Berlin: Walter de Gruyter), 193-215.
Wall, David (2012). "A Moorean paradox of desire." *Philosophical Explorations* 15.1: 63-84.
Wilde, Oscar (1893). "A woman of no importance." Oscar Wilde online. http://www.wilde-online.info/a-woman-of-no-importance.html. Accessed May 1, 2017.
Williams, Bernard (1976). "Deciding to believe." In Bernard Williams, *Problems of the Self: Philosophical Papers 1956-1972* (pp.136-151). Cambridge: Cambridge University Press.
Williams, John N. (1996). "Moorean absurdities and the nature of assertion." *Australasian Journal of Philosophy* 74: 135-49.
Williams, John N. (1998). "Wittgensteinian accounts of Moorean absurdity." *Philosophical Studies* 92.3: 283-306.
Williams, John N. (2004). "Moore's paradoxes, Evans's principle and self-knowledge." *Analysis* 64.4: 348–353.
Williams, John N. (2014). "Moore's paradox in belief and desire." *Acta Analytica* 29: 1-23.
Wilson, George M. (2004). "Comments on 'Authority and Estrangement'." *Philosophy*

and Phenomenological Research 69.2: 440-447.
Wittgenstein, Ludwig (1953/2009). *Philosophical Investigations*. Trans. G.E.M. Anscombe, P.M.S. Hacker and Joachim Schulte. Oxford: Wiley Blackwell.
Wittgenstein, Ludwig (1980). *Remarks on the Philosophy of Psychology* ii, ed. G. von Wright and H. Hyman. Chicago: University of Chicago Press.
Wittgenstein, Ludwig (1958). *The Blue and Brown Books*. Oxford: Blackwell.
Wood, James (2008). *How Fiction Works*. New York: Farrar, Straus, and Giroux.
Wright, Crispin (1989). "Wittgenstein's later philosophy of mind: Sensation, privacy, and intention." *The Journal of Philosophy* 86.11: 622-634.
Wright, Crispin (1998). "Self-knowledge: The Wittgensteinian legacy." In *Knowing our Own Minds*, ed. Crispin Wright, Barry C. Smith, and Cynthia MacDonald (Oxford: Clarendon Press), 13-46.
Yalcin, Seth (2007). "Epistemic modals." *Mind* 116: 983-1026.
Zagzebski, Linda (1994). "The inescapability of Gettier problems." *The Philosophical Quarterly* 44.174: 65-73.