UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Agency and First-Person Authority

Permalink
https://escholarship.org/uc/item/2rb7s823

Author
Parrott, Matthew Thomas

Publication Date
2011

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Agency and First-Person Authority

By

Matthew Thomas Parrott

A dissertation submitted in partial satisfaction of the

Requirements of the degree of

Doctor of Philosophy

in

Philosophy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Barry Stroud, Co-Chair
Professor Daniel Warren, Co-Chair
Professor Alison Gopnik

Spring 2011

© 2011
Copyright
Matthew Thomas Parrott
All rights reserved

Abstract

Agency and First-Person Authority

By

Matthew Thomas Parrott

Doctor of Philosophy in Philosophy

University of California, Berkeley

Professor Barry Stroud, Co-Chair
Professor Daniel Warren, Co-Chair

Ordinarily when someone tells us about her psychological states, we presume that she is right. By deferring to her in this way, we treat her as a kind of authority on her own psychological life. Although a person usually has this authority, she lacks it whenever she takes a more detached, indirect, or third-personal point of view toward her psychological states. We see this, for example, when she learns about a belief or desire from a friend or therapist. For this reason an adequate account of the phenomenon of "first-person authority" must explain why we have it only for some but not all of our psychological states. Most philosophers believe first-person authority is an epistemic phenomenon, consisting in each of us being better situated to know about our own psychological states than anyone else. Against all such epistemic views, I argue that, because they base their accounts on epistemic privileges that are in principle available to anyone, they cannot capture the exclusively first-personal character of our authority.

As an alternative to the traditional approach, I argue that first-person authority is derived from a person's agency with respect to her own psychological states. By relating to her psychological states in a first-personal way, a person is able to change or maintain them directly on the basis of what she takes to be good reasons for them. Since no other person can affect her psychological states in this way, her capacities as an agent guarantee her a unique kind of authority for them. A person ordinarily expresses this kind of agential authority over her psychological states in what she says about them. This is what justifies our deferring to her psychological self-ascriptions. On the view I develop in this dissertation, first-person authority is not primarily a matter of special epistemic access to psychological facts and deference is not a response to the epistemic status of what someone says. It is an acknowledgment of the special role that a person's agency plays in determining her psychological life.

For Daniel, in loving memory

Table of Contents

Chapter One: What is First-Person Authority?	1
Chapter Two: Senses of First-Person Authority	12
Chapter Three: On Deferring	37
Chapter Four: Self-Blindness and Self-Awareness	59
Chapter Five: Introspection, Illusion, and Consciousness	79
Bibliography	102

ACKNOWLEDGEMENTS

This dissertation owes its existence to many people. First of all, I am grateful to the University of California's Department of Philosophy, which provided a wonderful intellectual environment in which to study philosophy. While writing this, I was very fortunate to have been surrounded by very thoughtful individuals whose ideas, comments, and questions substantially improved my own thinking on the topic of this dissertation. Specifically, I thank Joseph Barnes, Josh Beattie, Justin Bledin, Zack Bruce, Ben Callard, Fabrizio Cariani, Stanley Chen, Vanessa De Harven, Kenny Easwaran, Nick Jones, Markus Kohl, Niko Kolodny, Tamar Lando, Berislav Marusic, Jennifer Marusic, Ian Phillips, Michael Rieppel, Stephen Schmall, Ian Schnee, John Schwenkler, James Stazicker, and Mike Titelbaum.

Over the years, I have learned tremendously from the philosophy faculty at Berkeley. I would like to thank each of them for being supportive and especially stimulating philosophers. John Campbell, Geoff Lee, Niko Kolodny and Mike Martin deserve special thanks. Each provided extensive and detailed feedback on multiple aspects of this dissertation and many hours of enlightening conversation. Outside the philosophy department, I must give a special thanks to Alison Gopnik who offered much helpful advice, criticism, and encouragement on this project.

Barry Stroud and Daniel Warren have been exceptional teachers. Their persistent questioning of assumptions and tireless pursuit of philosophical understanding compelled me to think harder and more carefully about the topic of this dissertation than I thought was possible. Each of them, in his own way, has provided a model of the kind of philosopher I hope to become.

I most grateful to four people: Andreas Anagnostopoulos, Tony Bezsylko, David Ebrey, and Josh Sheptow. The majority of my work on this dissertation consisted of thinking through and responding to their various ideas. Our many hours of discussion inspired most of the thoughts on these pages. Whatever insights I have managed to convey are partly theirs. They have been the best colleagues anyone could want and wonderful friends.

Finally, I would like to thank my friends and family for all their support and encouragement. Thank you to my parents for always believing in me and to my brother for commiserating about the process of writing a dissertation. Finally, thank you to Beth for standing by my side, sustaining me through all the hard times, and, most importantly, for filling my life with love and happiness--it truly does make everything better.

CHAPTER 1
WHAT IS FIRST-PERSON AUTHORITY?

If you want to find out what someone thinks about the weather, wants to eat for dinner, or fears most of all, asking her seems like the best thing to do. All of us presume that what someone says about her own psychological states is true; we defer to what a normal person says about the existence and character of her own beliefs, desires, feelings and moods. For instance, if I were to sincerely tell you that I believe that Cal has a much better football team than Stanford, you would typically take what I say to be true. Based entirely on what I say about what I believe, you may very well come to believe that I believe that Cal has a better team than Stanford, regardless of what you think about either team. Most of the time, you do not even consider the possibility of other evidence indicating what my belief is. Rather, from your point of view it seems that what I *say* about what I believe conclusively settles the question of what it is that I do believe. This is how you would treat me even when I am a perfect stranger to you. So, it does not seem that you must have any kind of background knowledge about how well I keep track of my own beliefs or desires or that you need to know whether my prior declarations about what I believe have been consistently true in the past. Independent of these considerations, my saying something about my beliefs, desires or feelings is sufficient for you to discern what I do in fact believe. This phenomenon of deference is puzzling, however, because what I believe is just a contingent fact, a fact that could easily have been different, and what I *say* about any other kind of contingent fact in the world is not presumed to be true in this way.

This puzzling phenomenon is widespread. What any person *says* about what he or she believes, desires, feels, or intends is, normally, presumed to be true. This practice of deferring to a person's self-ascriptions of psychological states is a distinctive way that we treat an individual as a kind of authority with respect to his or her own psychological life. By deferring in this way, we treat her statements as the final word on what her psychological states actually are. We therefore seem to be responding to some kind of special authority the person has in virtue of having or owning the psychological state she is speaking about. She seems to stand in a distinctive authoritative relation to only her own psychological states.

On closer inspection, however, we can see that simply owning a psychological state is not what secures this authority. At times, a person can have a psychological state from which she is alienated or detached; a state that, strictly speaking, belongs to her but does no cohere with the central aspects of the rest of her psychological life. For instance, she may have an uncomfortable desire repressed years ago, or an odd belief about a friend or loved one that sits uncomfortably with everything else she thinks about the individual. These sorts of cases should be familiar to most adult humans. It seems clear that this kind of relationship to a psychological state is fundamentally different from the kind a person has to

psychological states she wholeheartedly endorses as her own, as states central to her self-conception. In fact, the relation is quite similar to the kind of third-personal relation a person stands in to the psychological states of other people. Indeed, sometimes a psychological state can feel so foreign that it may as well belong to some other person. Whenever a person takes up this sort of third-personal point of view on her own psychological life, whenever she relates to one of her own beliefs, desires, or feelings in the same kind of way she would relate to another person's, she does not seem to stand in authoritative relation to it. It is for this reason that we do not defer to what a person says about beliefs or desires she learns about only after a lengthy process of psychotherapy or by making inferences based on observations of her own behavior. The third-person perspective, whether on another person's psychological life or on one's own, carries no intrinsic authority.

So, it seems that the authority we defer to by presuming a person's psychological self-ascriptions are true is an exclusive feature of the first-person point of view, to the point of view a person adopts in virtue of being the subject engaged with her psychological states. Only when a person relates to her beliefs, desires, intentions and feelings in an engaged, first-personal way is she an authority on their existence and character. This is at least how things have struck most philosophers writing about this phenomenon, which they have therefore come to call "first-person authority". For example, Jane Heal writes in a paper entitled "On First-Person Authority" that "what people say, in the first-person and present tense about their own thoughts is treated as authoritative."[1] And Barry Smith claims that the reason that others "do not ask us for justification," for what we say about our own minds, the reason that they defer to our psychological self-ascriptions, is that "they regard us as authorities on matters of our own psychology."[2] Everyday considerations like deference suggest the presence of some distinctive kind of authority intrinsic to the first-person point of view. But what exactly is this first-person authority? What phenomenon are we picking up on when we defer to a person's psychological self-ascriptions? How does our having first-person authority entitle others to defer to what we say about our beliefs, desires, and feelings?

This dissertation attempts to answer these questions. My primary aim is to understand the nature of the special kind of authority each of us seems to possess for only our own psychological states and only when we relate to them in an engaged first-personal way. This is crucial for understanding why other people are reasonable to presume that what we say about our own psychological states is true. Answering these questions, however, will also show us ways in which a person's ordinary first-personal relation to her own psychological life is fundamentally different from the way she relates to the psychological lives of others. It should not be surprising that each of us stands in a special relation to our own beliefs, desires, and feelings. Thus,

---

[1] Heal (2001).

[2] (1998): pg. 393.

by accurately understanding what is authoritative about this relation, I hope to also shed considerable light on what precisely is distinctive and special about the first-person point of view.

I

Throughout the history of philosophy, first-person authority has been understood in purely epistemic terms. According to the traditional line of thinking, first-person authority is a kind of epistemic authority, derived from some kind of epistemic privilege or advantage that each person has with respect to her psychological states and only her psychological states. It is, on this view, because each of is in a better position than anyone else to know facts about our own psychological lives that we are authorities. Each of us is an authority by being a kind of epistemic expert on the domain of psychological facts that are our own.

This interpretation of first-person authority is usually traced back to Descartes. Descartes believed that our access to our own psychological states was epistemically more secure that our access to any other kind of fact. Thus, he argues in the *Meditations*, that even though he may be deceived about the existence of external world, including the minds of others, it is impossible that he is deceived about the existence of his own psychological states.[3] If this were right, it would seem to be impossible for us to be wrong about our own psychological states and our ordinary way of knowing about them would be far more accurate and much less likely to be wrong than any ways of knowing available to other people. We would stand in a unique epistemically privileged position that allows us assess their existence and character, making us authorities on them.

Most contemporary philosophers reject Descartes' account of the nature of psychological states and how we come to know them. Moreover, as Freud made vividly clear, there is no good reason to think that we are infallible about our own psychological lives. Nevertheless, philosophers continue to assume that first-person authority is to be understood in an epistemic sense and that an account of it should basically be trying to explain a fundamentally epistemic phenomenon--why each of us is stands in an epistemically privileged position for knowing about only our own psychological states. Thus, ever since Descartes philosophers have offered many alternative accounts of the nature of the epistemic privilege a person possess in virtue of taking up the first-person point of view and most recent writing on first-person authority consists of disputes between contemporary philosophers on how

[3] Although the idea that each of us is infallible about the character of our own psychological states is usually attributed to Descartes, it is actually not obvious that Descartes thought this. Relevant subtleties of Descartes' own view are discussed in Broughton (2008) and Carriero (2009).

best to explain this first-person epistemic privilege. But, instead of entering these debates, I think we should first notice that there are alternative, non-epistemic, ways of understanding the concept of authority.

In addition to the epistemic kind of authority a person can have in virtue of being in a privileged position to know some fact or other, there is also the type of authority a person can have by being in a better position to do something or other. I will call this latter type of authority agential authority. Whereas a person has epistemic authority by being in the best position to know what some fact is, a person has agential authority by being in a better place to determine what some fact is, to constitute the relevant fact. For example, my mechanic is an epistemic authority when it comes to facts about my car engine. He is in the best position to know what is or is not the case. But a military general is an authority when it comes to the location of his troops; but not because he is an epistemic expert. He may not even be in the best position to know where his troops are; that person might be their field commander. But the general is an authority nevertheless because he is in the best position to determine the location of his troops by ordering them to that location. Other things being equal, what the general decides determines where his troops go. The general is therefore a kind of agential authority; he is an authority in virtue of his capacity to act in a special way. I will return to the difference between these two kinds of authority in the second chapter, but I hope these brief examples show that we can make a distinction between different senses of authority, a distinct that bears directly on understanding the phenomenon of first-person authority.

In this dissertation I shall argue that first-person authority is a type of agential authority; it is a special kind of authority each of us has in virtue of specific features of our cognitive agency. Because we are agents with respect to our own psychological states, we have the ability to directly determine their existence and their character. This is something no other person can ever do. For this reason, our capacity to function as cognitive agents guarantees us a special kind of authority on our own psychological lives.

My attempt to ground first-person authority in cognitive agency will strike many philosophers as misguided because they will take it for granted that first-person authority is a kind of epistemic authority; they will assume it is the kind of authority we have in virtue of some epistemic privilege. But it is crucial to notice that this is substantial assumption. As I shall demonstrate, nothing about the phenomenon of our authority or about any of the more intuitive considerations suggesting we have a special kind of authority requires that we interpret it in epistemic terms. In fact, I shall argue that an epistemic account is ultimately *unable* to capture the phenomenon of first-person authority that we are responding to when we practice deference.

II

We can begin to see the shortcomings of the epistemic approach by first noticing that we do not always have first-person authority. In addition to the ordinary, first-personal way that I relate to my own psychological states, I can, at times, stand in a more third-personal relation to them. It is by now a familiar idea that we all sometimes learn about our own beliefs or desires by talking with our closes friends or family members. Alternatively, we may learn a great deal about what we really feel or want from our therapist. When I relate to one of my psychological states in any of these ways, however, it no longer seems to be a state for which I am in any way authoritative. It seems that I lose first-person authority when I do not relate to my psychological states in an engaged first-personal way.

So, although every person usually has first-person authority, she will lack it when she takes a more detached, indirect or third-personal point of view on her own psychological states. This suggests that having first-person authority depends on the way in which a person is related to her psychological states. It is not enough to simply own a psychological state; rather, one must stand in the appropriate first-personal relation to a psychological state in order to have authority with respect to it. *Prima facie,* there is something about the very nature of the first-person, about taking up the first-person standpoint, which is fundamentally responsible for this special kind of authority. This is another crucial point to notice because it means that an account of first-person authority needs to explain why we have it only for some and not all of our psychological states; why we have it only when we are engaged with them in a first-personal way.

Any philosophical account of first-person authority will therefore have to meet two conditions in order to be sufficient. First, it will have to explain a type of authority that is *essentially* first-personal. If this authority depends upon the way we are related to our own psychological states, it seems to be an intrinsic feature of the first-person. Therefore, what we are explaining cannot be a kind of authority that someone could possibly have when she relates to her psychological states in a more third-personal or detached manner. It also cannot be a kind of authority that she could have with respect to psychological states belonging to some other person. I want to stress that this is not a condition on the kind of explanation one develops; it specifies precisely how we must understand the *explandum*. The phenomenon of authority we are trying to account for is exclusive to the first-person point of view. This is why it has seemed to so many philosophers to be a special *kind* of authority.

The second condition on an adequate account is that it must ultimately explain how our authority grounds the practice of deferring to psychological self-ascriptions. Heal and Smith are not the only people who take deference to be clear evidence for first-person authority.[4] If we are authorities, it is not in some abstract

[4]Others include Davidson (1987), Wright (1998), and Fricker (1998).

way, but in a peculiar way that justifies other people taking what we say about our own psychological states to be true. Importantly, this practice of deferring is also peculiar; it not at like the kind of deference we extend to medical, scientific, or automotive experts. Unlike the practice of deferring to epistemic expertise, deferring to psychological self-ascriptions is immediate and seems to be warranted for any sincere speaker. It is not something that we practice when a person's psychological self-ascriptions are made from a more third-personal perspective, even though a person may very well become an epistemic expert in learning about her own psychological states from such a perspective. If, for example, I were to report that I feel rage toward my younger brother because of a conversation that I had with my therapist or because I noticed I was exhibiting hostile behavior towards him at our last family get together; in other words if I *tell* you that I am basing my self-ascription on evidence, my self-ascriptions no longer seem to warrant deference. Even if I am an expert on psychoanalytic interpretation and my self-ascription is based on the best possible evidence, it does not seem that you should immediately defer to what I say about my rage. These considerations suggest that deferring to psychological self-ascriptions is a unique sort of practice. Since it seems that we are picking up on a person's authority when we defer in this special way, a sufficient account of first-person authority should help to make clear why this is so.

These two conditions of adequacy help us focus on the real phenomenon of first-person authority. Neither one, on its face, requires that we interpret the authority in epistemic terms. This is significant to keep in mind because assuming a particular interpretation of the authority limits the possibilities of how we might best account for the phenomenon. In this dissertation, I will take these two conditions very seriously and develop a new account of first-person authority that meets them both. I shall argue that only an account of first-person authority in terms of cognitive agency can capture a distinctive kind of authority that is both exclusive to the first-person and capable of justifying our practice of deferring to psychological self-ascriptions.

If my account is correct, it means that the existence of an epistemic privilege cannot adequately explain first-person authority. But isn't it true that each of us *just does* know her own psychological states in a special way not available to others? Consider how in order to know what you believe I must observe you, either directly or indirectly, in some way or other. I must either see what you are doing, or listen to what you or someone else is saying, in order to know what your psychological states are. But, in my own case, none of this is necessary. When it comes to my psychological states, it seems that, in normal circumstances, I *just know* what they are; I don't have to investigate, deliberate, inquire or perceive anything. It therefore seems clear that I have a different kind of epistemic access to my own psychological life. Having this special first-person access motivates philosophers to pursue a more purely epistemic account of first-person authority.

Yet, the fact that we know about the existence and character of our own psychological states in a distinctive *way* does not by itself entail that this way is epistemically privileged. It does not mean that the first-person way of accessing psychological facts is epistemically *superior* to any of the other ways of knowing about them. A distinct kind of access is not equivalent to, nor need it be taken to be, an epistemically privileged mode of access--uniqueness is not the same thing as privilege. Thus it is possible for us to acknowledge a distinct kind of first-person access without having to think that we are each epistemic experts on our own psychological lives. Our first-personal way of relating to our own psychological lives is indeed special and has many distinctive features, but in this dissertation I will illustrate how we can best capture this way of having psychological states in agential terms.

III

In Chapter 2, I begin by discussing the two relevant senses of first-person authority, the epistemic sense and the agential sense. I argue that when first-person authority is interpreted in an epistemic sense, it is a type of authority that is not essentially first-personal. There is nothing about enjoying an epistemic privilege of any kind that requires a person be related to her psychological states in an essentially first-personal way. This means that traditional accounts of first-person authority in terms of an epistemic privilege fail to capture a kind of authority exclusive to the first-person; they fail to meet the first condition of adequacy. Regardless of how one explains the details of the supposed epistemic privilege, it will not be an essentially first-personal phenomenon; it will not be a king of authority grounded in the intrinsic features of the first-person point of view.

As an alternative, I offer my own interpretation of first-person authority in terms of agency and argue that one is an authority in virtue of being able to directly determine what one's psychological states are directly on the basis of reasons. Having this capacity does depend on being related to one's own psychological states in an engaged first-personal way. Only when we stand in a first-personal relation to one of our psychological states can we function as cognitive agents to directly determine its existence and character. Thus, my agential account succeeds in capturing a distinct kind of authority exclusive to the first-person; it meets the first condition of adequacy.

In Chapter 3 I focus on the second condition of adequacy. Initially, it might appear that nothing about agency, about merely being able to do something with respect to one's own psychological states could possibly account for why other people are justified in presuming any of our self-ascriptions are true. The very practice of deferring to what someone says is a way of conferring a special epistemic status on a class of assertions; deference therefore appears to be an epistemic

practice. It would make most sense if it rested securely on the epistemic status of the propositions we are presuming to be true, on a person's sincere reports about the existence and character of her own psychological states. It is difficult to see how agency alone could possibly explain this practice.

Nonetheless, I shall defend the view that deference to a person's psychological self-ascriptions is justified by a speaker's agential authority. Unlike the deference we practice toward epistemic experts in different areas like science or medicine, the practice of deferring to psychological self-ascriptions is completely insensitive to important epistemic parameters surrounding what a person says. To take just one example, a doctor or scientist may strengthen her entitlement to deference by appealing to supporting evidence, but a direct appeal to evidence in support of a psychological self-ascription actually undermines one's entitlement to deference. This is one fact that indicates that our justification for deferring to psychological self-ascriptions is independent from the epistemic context of the self-ascription. Because the surrounding epistemic context seems to be irrelevant, I argue that deferring to psychological self-ascriptions is not a response to the epistemic standing of a person's assertions. What then is it?

When a person with the agential authority to determine what her psychological states are says something about those states, I believe she expresses her authority over them. I shall argue that in deferring listeners are responding to this expression of authority. Deferring to what someone says about her own psychological states is our way of acknowledging that she alone has the authority to directly determine the existence and character of those states. Thus, although deferring is an epistemic practice on the part of the listener--it is a way of conferring a certain epistemic status on a class of propositions--it is not a practice we are justified in because of the epistemic properties of a speaker's assertion. If this explanation is correct, my agential account meets both conditions of adequacy for successfully understanding the phenomenon first-person authority.

Even if this is true, the notion of epistemic privilege is firmly entrenched in philosophical discussions of the first-person.[5] To many philosophers, it seems obvious that the ordinary knowledge a person has of her own psychological life is more secure than any other kind of knowledge could ever be. It will therefore be objected that a theory of first-person authority in agential terms misses a basic fact about first-person access, the fact that it is an epistemically privileged way of knowing. In Chapters 4 and 5, I discuss this general line of thought.

It is true that first-person access affords us a distinctive way of knowing about our own psychological lives. We ordinarily have a kind of access that is non-

[5] It is interesting in itself that this prejudice seems to only be widespread among philosophers. Most non-philosophers that I have spoken with over the years tend not to believe that each of us stands in an epistemically privileged relation to our own psychological states.

evidential and non-observational and that is enough to distinguish it from other ways of accessing the same psychological facts. But this alone does not entail that first-person access is epistemically privileged or that it is a better more secure way of knowing. Philosophers tend to think otherwise, I believe, because they hold one of two assumptions about first-person access.

The first assumption is that the nature of psychological states is such that a person could not possibly have a psychological state without knowing or being aware that she does. A version of this idea was made explicit by Locke who claimed that consciousness was "inseperable from thinking, and it seems to me essential to it: It being impossible for any one to perceive, without perceiving, that he does perceive."[6] According to Locke if you have any kind of psychological state, you also had awareness that you had it. The two are not distinct or independent entities. Rather the existence of a psychological state metaphysically guarantees knowledge of its existence.

Since Locke, many philosophers have been tempted by the idea that a psychological state is constitutively connected to its owner's knowledge of it in some form. Against this idea, in Chapter 4, I argue in favor of what I call the *Distinct Existence Thesis*. This thesis embodies a very natural idea that a psychological state is one thing and knowledge or awareness of it is something else. The two are ontologically independent, or as Hume would have said, they are distinct existences. I argue for the thesis on the ground that we need it to make sense of our fallibility about our own psychological states. If the *Distinct Existence Thesis* were false, we should not be as susceptible to mistakes or ignorance as we in fact are. For this reason, philosophers who deny the *Distinct Existence Thesis* owe us an account of how we make mistakes in this domain whenever we do. But, I argue that unless the *Distinct Existence Thesis* is true, there is no plausible explanation for why we are even sometimes mistaken about our own thoughts and feelings is available.

Accepting the *Distinct Existence Thesis* may initially seem to assimilate first-person access to other ways of knowing. If my psychological state is ontologically independent from my knowledge of it, it must in some way cause my knowledge of it; first-person access will therefore rest on some kind of causal relation. But our perceptual access to facts in the world rests on a similar kind of causal relation; material objects causally interact with our sense organs in such a way that, in favorable conditions, produces knowledge. If first-person access is analogous to perceptual access in this way, it may no longer seem to be a very distinctive way of knowing.

Because we have strong intuitions that first-person access is distinct from other ways of knowing, it is important for me to explain how it is. Earlier, I pointed out that our first-personal way of knowing is unique in being non-evidential and non-observational. In Chapter 4, I argue that, as a way of knowing, it not explained

[6] *Essay,* BkII.27.9

by any causal relations that hold between psychological states. Instead, I argue that a person is entitled *a priori* to know about her own psychological states in a distinctively first-personal way because she is fundamentally a cognitive agent in relation to her own psychological life. Because we could not cognitively act without being able to know about our actions in a special first-personal way, our cognitive agency explains why we are a priori entitled to take the contents of first-person access at face value. Thus, although the way we access our own psychological states may be realized by a causal process in the brain, the epistemology of first-person access need not be causally explained.

A second assumption also motivates philosophers to think our first-personal way of knowing is epistemically privileged. Those who make this assumption also focus on the fact that we require perception in order to know about another person's psychological states. Because of this, they assume that our first-personal way of knowing psychological states is not as susceptible to errors due to failures in the perceptual system. If this were true, then our first-personal way of knowing psychological states would in fact be more reliable and less susceptible to error than any way of knowing available to others. Empirically, it would *be* epistemically privileged even if only for contingent reasons.

In Chapter 5 I argue against the notion that our first-personal way of knowing is epistemically privileged. First, I examine some everyday cases where people are mistaken about their own thoughts and feelings. The cases seem to be commonplace and occur with no less frequency than mistakes about facts in the external world. This is important because if each of us were truly enjoying an epistemic privilege on our own psychological states, one would expect these cases to be rare, which they are not. It therefore does not seem that the first-personal way of knowing is epistemically privileged.

In further support of this claim, there has been a tremendous amount of recent research in empirical psychology that clearly indicates we do not in fact enjoy an epistemic privilege when it comes to knowing about our own psychological states. Since roughly 1980, social psychologists have conducted an impressive range of experiments that show people are often mistaken about their beliefs, feelings, passing thoughts, and even their own experiences. The data is compelling and I believe it is sufficient to show that the first-personal way of knowing is not epistemically privileged in any sense. Although we are not yet aware of precisely what physical processes realize first-person access, evidence indicates that they are just as susceptible to breakdown and misfiring as the ones that realize perceptual access. Since the first-personal way of knowing is no better than perceptual ways of knowing, I conclude that there is no epistemic privilege connected to the first-person point of view. There is therefore no reason to think that we must account for why we have some kind of epistemic advantage over others in knowing what is happening inside our own minds.

Anyone can wonder why we seem to be authoritative with respect to our own psychological states. Noticing the phenomenon of deference may easily push us toward thinking of our relationship to our own mind in terms of epistemic privileges. But we can also notice that other people are often in a better position to know what we are thinking or feeling. Sometimes our close friends or family members are better at discerning what we really want, think, of feel. How does that happen? If we notice how often we are wrong about ourselves, it suggests that we do not have an epistemically privileged point of view on our own minds. We are then faced with a puzzle. Why would people defer to what I say about my own beliefs or desires if not because I am in an especially good position to know what they are? If other people know as well as I do what I think or feel, why should they just listen to what I say?

I am convinced that the right way to untangle this puzzle is to notice that cognitive agency is fundamental to our first-personal way of having psychological states. From the first-person point of view, I am in a position to directly adjust, change, and act on my own beliefs, desires, and feelings. For example, I can change my beliefs simply by judging that I have reasons for believing things. These kinds of cognitive actions are essential to the way we understand persons as subjects of psychological states. They are, more importantly, exclusive to the first-person--no one else can affect my psychological life as directly or in the same ways. It therefore strikes me as strange that agential aspects of our mental lives have not been even *more* prominent in philosophical discussions of the first-person. In the following chapters, I focus a great deal on ways in which we are cognitive agents. This is because I think that coming to a better understanding of our cognitive agency helps very much with our understanding of ourselves and our place in a world of other, like-minded, persons.

CHAPTER 2
SENSES OF FIRST-PERSON AUTHORITY

In general, the concept of "authority" has different senses in our language. Sometimes a person has authority in an epistemic sense, by being better situated than anyone else to assess evidence, to make relevant observations or to report on certain facts. The leading authorities in biology or medicine are epistemic authorities; they either know more or are better placed than others to learn about facts in each of their respective fields. But a person may also be an authority by being better situated than other people to do something, to exert control over something, to be responsible for something, or to determine something in a unique way. Consider an umpire at a ballgame. The umpire is an authority for whether the pitch just thrown is a strike. Similarly, I am an authority for the grades on student papers in the philosophy courses I teach. Although each of us may in fact know better than most people what strikes or passing grades are, our knowledge is not what secures our authority. Rather, the umpire and I both have authority in a different sense, what I will call an agential sense. We are authorities because we are the only ones responsible for determining the relevant facts in a given domain.[7] Everyday considerations like deference reveal the presence of a unique first-person authority, but what sense of "authority" best captures the phenomenon we are noticing?

Philosophers traditionally understand first-person authority in a purely epistemic sense, taking it to consist in each person being in a better position to know about the existence and character of her own psychological states. There are two main problems with this approach to understanding the phenomenon. First, there is much evidence from recent psychological research that seems to indicate each of us does not, in fact, enjoy any epistemic advantage over others when it comes knowing about our own psychological states. If things are as this research indicates, there is no first-person authority in the epistemic sense.[8] Although this is a serious difficulty for an epistemic interpretation of first-person authority, in this chapter I want to focus on the second problem.

The phenomenon of first-person authority seems to be exclusive to the first-person standpoint; it seems to be a special kind of authority, something essentially tied to the first-person. But an epistemic understanding of our authority is not faithful to this first-personal character. It accounts for our authority on the basis of epistemic privileges that could in principle extend to any psychological state, including those toward which we take a detached third-personal perspective. If we

---

[7] I do not intend these two senses to be exhaustive. There are other senses of "authority" but I do not think they are relevant to the topic of this chapter.

[8] This body of research is extensive and seems to grow every year. For a very brief overview of some recent research, see Wilson (2002).

try to understand first-person authority in an epistemic sense, it will not require that a person be related to her psychological states in an engaged first-personal way in order to possess authority. Therefore, nothing about being an epistemic authority or having an epistemic privilege could sufficiently explain why the first-person point of view has a special *kind* of authority.

As an alternative to the traditional epistemic approach, I believe that we should understand first-person authority in a purely agential sense. An account of first-person authority in terms of agency holds that when a person is related to her own psychological states in the ordinary first-personal way she is better situated to *do* something with respect to them. More specifically, I shall argue that a uniquely first-personal relation to one's psychological states endows a person with a special capacity to determine what her own psychological states are directly on the basis of her own sense of good reasons for them. Unlike epistemic accounts, this agential account does explain a kind of authority that requires the most central aspects of our first-personal way of relating to our own psychological states.

I

First-person authority is never present when we take up a third-personal perspective on the beliefs of other people. As we know, we can also take such a point of view on ourselves and, when we do, we seem to lack first-person authority. Whenever we learn about our psychological states in these indirect ways, we self-ascribe them on the basis of observable, public, evidence, just as when we ascribe psychological states to other people.

Suppose you were to ask me whether I believed that Berkeley is a nice place to live, I could investigate my personal history in order to answer your question. I do participate in an above average amount of activities in Berkeley. I also frequently tell my friends and family nice things about Berkeley, much more than about Detroit where I grew up. It might be clear to me on the basis of all the observable evidence that I do indeed believe that Berkeley is a nice place to live. In cases like this, however, my pronouncements about my own psychological states are not entitled to the deference of others. If you knew that I learned about my belief by observing my own behavior, you would not immediately presume that what I said about it was true. That is because the third-personal perspective on one's own psychological life is not an authoritative one--it is the same kind of perspective any other person can take on my psychological states. Shifting into this perspective on one's own beliefs is therefore *like* treating your own beliefs as if they were anyone's.

In order to make an authoritative psychological self-ascription, one that warrants the deference of others, it seems that a person must be related to her psychological states in an engaged first-personal way rather than a detached third-personal way. There are various other differences between these two ways of having

a psychological state. I think that focusing on asymmetries between the first- and third-person can help us better understand what kind of authority is attached to the first-person point of view. In what follows, I will primarily discuss the psychological state of believing because this allows me to engage more directly with recent work in philosophy. But, analogous points can be made about other types of psychological states.[9] Comparing the relationship I normally have to my own beliefs to the third-personal ones I can have to another person's, reveals characteristic features of the first-personal way of having a psychological state, some of which seem especially relevant to understanding the phenomenon of first-person authority.[10]

One obvious difference is that in my own, first-personal case, my relationship to my beliefs rarely involves any reflection on or observation of my beliefs. As a person, I am most often focused on the things around me: my friends and loved ones; the best ingredients for tonight's dinner; or difficult philosophical passages. My attention and my beliefs are both typically directed at facts in the world outside of me. In this way, a person's beliefs and other psychological attitudes make up her perspective on the world. This is not true when it comes to my relationship to someone else's beliefs. I have to stand in a third-personal relation to your beliefs. I must take up some observational perspective or point of view in order to learn about them. By contrast, most of the time, I am related to my beliefs simply by *believing* things, by being a believer.

Yet even when a person does self-consciously reflect on her beliefs (because she is asked about them, for instance), she normally continues to direct her attention out at the world, to qualities of her friend or to her dining options for the evening. When you ask me what I *believe* about living in Berkeley, I will ordinarily, though not always, answer by considering appealing features of Berkeley: the high quality of local

---

[9] As examples, Moran (2001) and Bilgrami (2006) both offer quite elaborate views of first-person authority that focus on beliefs. Analogous accounts can, I think, be given for other types of rationally sensitive psychological states. These are what Scanlon calls "judgment-sensitive attitudes" (1998) and what Hieronymi (2005, 2008) calls "commitment-constituted attitudes". I am, however, doubtful the same is true for sensations. Sensations seem to be a distinct from psychological states like belief in a number of ways that will likely matter a great deal to the question of first-person authority. Three differences strike me as most relevant: 1) sensations are passive psychological phenomena; 2) sensations are not, even indirectly, determinable by a consideration of justifying reasons and 3) sensations have more of an immediate event-like quality than psychological states like belief that tend to persist (in other words, sensations naturally tend to end while beliefs tend to endure). I will set the topic of sensations aside for the rest of this chapter.

[10] Other characteristic features seem to be less relevant. One that is often discussed is immunity to error through misidentification. Although it is very interesting, I do not think it helps us understand the kind of authority I am interested in.

restaurants; the accessibility of recreational activities; the temperate climate. Thus, even in cases where you most expect a person to turn her attention inward toward her own psychological states, she continues to attend to things in the world outside of her.

This outward directed aspect of our ordinary way of relating to our own beliefs was famously noted by Gareth Evans:

> In making a self-ascription of belief, one's eye's 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*. (1982, pg. 225)

If somebody asked *me* whether *you* believed that there will be a third world war, however, it would be wrong to consider only "the same outward phenomena" that bear on the question whether there will be a third world war. Reflection on your beliefs requires that I consider them as facts independent of what they are about. I must consider psychological evidence in order to determine the existence and character of your beliefs. Contrary to this, as Evans's passage rightly highlights, our ordinary relation to our own beliefs does not involve this sort of reflection.

Evans is not saying in this passage that when a person is asked a question about her own beliefs she must actually explore the world around her or consciously implement some deliberative process. Rather, he says that only "occasionally" do we "literally" direct our "eye" outside in order to answer questions about our beliefs. For all that Evans has said, it is perfectly reasonable to expect persons to instantaneously answer questions about some of their beliefs upon being asked about them. Evans's talk of directing our "eye" is a metaphor to show us that the considerations relevant to answering a question about one's own belief, as opposed to a question about another person's belief, are *external* facts not inner psychological ones. Sometimes we might have to actually "get ourselves into position" to answer these questions by deliberating but other times we will, so to speak, already be in the right position.[11]

It would be a mistake, however, to conclude from this that the reason we attend to what Evans calls "outward phenomena" is because it functions as evidence for our self-ascriptions. Rather, as Donald Davidson pointed out, "first person

[11] Cf. Martin (1998).

attributions are not based on better evidence but often on no evidence at all."[12] Psychological self-ascriptions, whether in speech or thought, are usually made independently of all evidential considerations. This importantly qualifies how we should understand Evans's talk of directing our attention outward.[13]

Davidson's point about independence from evidence is a purely negative point about the epistemology of our ordinary first-personal way of relating to our own beliefs; nothing about it entails that the first-personal way is epistemically privileged.[14] As he points out, self-ascriptions of belief do not normally have an epistemic basis, but this does not imply anything about the epistemic status of those self-ascriptions. In order to support a more robust epistemological thesis, one would have to explain in far more detail precisely what about this particular way of self-ascribing makes it epistemically privileged over ascriptions that are based on evidence. As Davidson recognized, "claims that are not based on evidence do not in general carry more authority than claims that are based on evidence."[15]

Both Davidson and Evans highlight distinctive features of the first-personal way we ordinarily relate to our own beliefs. Unlike relations we stand in to the beliefs of other people, each of us is usually related to our own by having beliefs focused on the facts in the world. As Evans notes, we normally maintain this kind of relation when questioned about our beliefs without having to step back to a more reflective

---

[12] Davidson (1984), pg. 6. This non-evidential feature of our psychological self-ascriptions is central to what Burge (1988) has in mind when he claims that "brute errors" are impossible when self-ascribing psychological states. An error is "brute" in Burge's sense when it occurs despite the person having a sound epistemic basis. The reason that such errors are impossible for psychological self-ascriptions is simply that there is no epistemic basis. Thus the non-evidential character that Davidson and Burge discuss is distinct from the fact that psychological self-ascriptions are non-inferential. The latter would be consistent with their having an epistemic or evidential ground, as, for example, one may plausibly think is true for perceptual judgments.

[13] It might also present a *prima facie* puzzle. If we do not consult "outward phenomenon" for evidence, then why do we do what Evans says? Why do we attend to these things in the external world? But this puzzle arises only if we think of Evans's point requires an epistemic justification. If, as I shall argue, there is another explanation, it should not be puzzling that these "outward phenomena" do not function as evidence.

[14] cf. Moran (2001). Peacocke (2008) makes a similar point concerning the distinction between action awareness and perceptual awareness, "The premise of the fallacious argument rightly alludes to the distinction between action awareness and perceptual awareness. This difference in kind does not by itself produce any kind of philosophically significant restriction on fallibility." (pg. 12)

[15] Davidson (1984), pg. 5

point of view on them. Moreover, since we can only ascribe beliefs to friends and colleagues on the basis of evidence that is available to anyone, the non-evidential character of our self-ascriptions is also a mark of the first-personal way of relating to one's own beliefs. Thus, directly engaging with our beliefs in a first-personal way requires that we adhere to the following two principles:

Evans Principle: Our attention is focused on the way things are in the world, not on our own beliefs independent from the facts they represent.

Davidson Principle: Our belief self-ascriptions are not based on evidential considerations; they are made independently of evidence.[16]

When a person violates either one, she takes up a more third-personal perspective on them. Since it seems clear that a person cannot continue to have first-person authority when she takes up third-personal point of view on any of her own psychological states, an adequate account of first-person authority should explain a kind of authority that requires a person to be in accord with these two principles. What I will argue in what follows is that this condition cannot be satisfied if we interpret the phenomenon of first-person authority as a kind of epistemic authority, if we take it to be something with an epistemic sense. One can possess epistemic authority for one's own beliefs even in cases where she violates the *Evans Principle* or the *Davidson Principle*.

II

When we relate to our beliefs in a more third-personal manner we lack first-person authority for them, but how? Consider the following case where a person's relation to her own beliefs is obviously third-personal in form, a case where she acquires knowledge of her beliefs through a process of psychotherapy.[17] Richard Moran calls this a "familiar therapeutic context" and he describes it as follows:

---

[16] Davidson presents his point as holding for all types of psychological states but Evans presents his insight as concerning only beliefs. Although I think that both of these principles hold generally for all types of psychological states, I restrict them to beliefs for the purposes of this chapter.

[17] Plausibly, the first- and third-personal way of relating to one's psychological life is a matter of degree. Thus, we should expect to find, as I think we do, than an individual can take more or less third-personal perspectives on her own psychological states.

> The person who feels anger at the dead parent for having abandoned her, or who feels betrayed or deprived of something by another child, may only know of this attitude through the eliciting and interpreting of evidence of various kinds. She might become thoroughly convinced, both from the constructions of the analyst, as well as from her own appreciation of the evidence, that this attitude must indeed be attributed to her. And yet, at the same time, when she reflects on the world-directed question itself, whether she has indeed been betrayed by this person, she may find that the answer is no or can't be settled one way or the other…She can only learn of it in a fully theoretical manner, taking an empirical stance toward herself as a particular psychological subject. [18]

Although I will be concentrating on an example that fits Moran's description, it would be a mistake to think that taking a more third-personal stance toward one's own beliefs requires psychoanalytic therapy. Quite often, we learn about what we believe through conversations with colleagues in our department, family members and close personal friends.[19] Moran's example is useful, however, because it vividly illustrates what happens when a person adopts a third-person standpoint on her own beliefs, which in turn helps us understand more clearly how she lacks first-person authority.

Let us call the person in Moran's example "Janet". Janet, we can presume, is suffering from many disquieting symptoms, ones that lead her to initially seek out psychotherapy. For my purposes, I will focus on only one problematic belief--the belief that her sibling betrayed her. At the start of therapy, Janet is completely unaware that she believes her sibling betrayed her. She would, if asked, consistently avoid self-ascribing it and she may even claim to have no opinion on the matter at all. What Moran has in mind is that over time a person like Janet can, through therapy, move to a second stage of her therapy and come to learn about the existence and character of her unconscious beliefs. We are to imagine a scenario unfolding over the course of therapy such that Janet eventually begins to self-ascribe the belief that her sibling betrayed her. At first, she may do so tentatively, not quite

---

[18] (2001), pg. 85. The passage is slightly misleading because Moran talks about attitudes like fear in the context of the analysis. But he uses the passage to bring out a truth about his "Transparency Condition", which is explicitly phrased in terms of belief. So, I'm going to assume that what is lying behind Moran's odd phrasing of this passage is a view about the cognitive conditions on feelings of betrayal; i.e., that in order to "feel betrayed" I must believe that someone betrayed me. Having the feeling entails having the belief. At any rate, since, like Moran, my discussion is focused on belief, I am going to assume this for the sake of simplicity.

[19] Bilgrami (2006) calls these ways of acquiring knowledge "cognitive forms of self-inquiry".

trusting her therapist. But Janet's therapist, rather than simply telling Janet what she really believes, coaxes her along in such a manner that she learns how to appreciate the evidence for herself. She notices, for instance, that she is extremely hostile toward her sibling during family holidays or that she feels anxious during their telephone conversations. Over months of therapy, Janet eventually learns what she believes and this allows her to accurately self-ascribe the belief that her sister betrayed her. What is most important to recognize is that when at this second stage Janet begins to self-ascribe the belief that her sibling betrayed her, she takes a third-personal perspective on that belief.

The characteristics of the distinctively first-personal way of relating to one's own beliefs are notably absent in Janet's scenario. Janet believes that she was betrayed by her sibling but her relation to the belief is exceedingly reflective; it is one that she acquires only after hours of detached reflection in therapy. She violates the *Evans Principle* because her attention is not directed outside at features of her sibling that ought to bear on the question of her betrayal. Even when she does consider facts about her sister, they do not seem to matter to what she believes; she must also consider the belief that her sister betrayed her as something independent from these facts. This is why, in Moran's words, "when she reflects on the world-directed question, whether she has indeed been betrayed by this person, she may find that the answer is no or can't be settled." Janet also violates the *Davidson Principle*. She bases her self-ascription on "her own appreciation" of psychological evidence. She correctly says that "I believe that my sibling betrayed me", but it is because she observes things like her own uneasiness on the phone or her hostility toward her sibling at Thanksgiving. In violating these two principles, Janet treats the belief that she was betrayed by her sibling third-personally, similar to how she would treat another person's beliefs.

Janet's case is important because it shows us that the first-personal features of a person's relation to her own beliefs cannot be adequately understood in epistemic terms. Although Janet takes a third-personal stance toward her belief in the second state of her therapy, she does not appear to be in any inferior epistemic position with respect to it. Through therapy Janet learns a fact about what she believes; she therefore knows something about herself-- she *knows that* she believes that her sibling betrayed her. Her knowledge of what she believes, *qua* knowledge, is completely stable; she has based it on very good evidence acquired through therapy. Moran describes Janet's problem as one where she "cannot learn of this attitude of hers by reflection on the object of the attitude." But, in claiming this he seems to give more credit to epistemic considerations than he ought to. We can see that to overcome her problem, Janet will have to move from the second stage of her therapy to a third stage where she stands in an engaged first-personal relation to her belief that her sibling betrayed her. But that will not really require Janet to *learn* something

about her attitude or acquire any more knowledge of it.[20] Janet already knows the facts about what she believes about her sibling before she begins to relate to her belief in a first-personal way.[21] If this characterization of Janet's situation is correct, then what she must develop in her therapy does not seem to be epistemic.

One way to see this last point more clearly, is to imagine Janet becoming a sort of psychoanalytic expert.[22] In that case, she could continually make reliably accurate self-ascriptions of many psychological states, but they would always violate the *Davidson Principle*. That would not obviously prohibit Janet's expertise from enabling her to be in a better position to know about her own beliefs than anyone else. And, this may well make Janet a kind of epistemic authority on her own psychological states. If we were solely interested in accumulating information about Janet's psychological life, she might be a reasonable place to go. However, any kind of authority that Janet could acquire by becoming an expert in this way would not be *first-person* authority because her way of relating to her beliefs would remain detached, distant, and third-personal. The ability to make reliable psychological self-ascriptions is therefore insufficient for having an authority that is exclusive to the first-person.

Rather than lacking something epistemic, Janet seems to lack a crucial ability to change her belief that she was betrayed by her sibling. This belief is completely insensitive to Janet's own sense of reasons for believing. If we were to ask Janet to reflect on facts about her sibling, to reflect on what Moran calls the "world-directed

---

[20] The same point was made by Freud (1911) who insisted that "If knowledge about the unconscious were as important for the patient as people inexperienced in psychoanalysis imagine, listening to lectures or reading books would be enough to cure him. Such measures, however, have as much influence on the symptoms of nervous illness as a distribution of menu-cards in a time of famine has upon hunger." (XI, 225) Freud's point here, as elsewhere, is that what must happen in order for psychoanalysis to succeed is not an increase in the patient's knowledge.

[21] Someone could object that Janet actually does acquire knowledge during therapy. On a certain conception of information, the perspective Janet takes or the relation she stands in to her belief actually changes the available facts. According to this line of thinking, shifting into a more first-personal relation on a psychological state is a way for Janet to learn something new because it creates a new fact rather than, as I have urged, being a way for her to relate in a different way to a fact that she has already learned. The same kind of issue arises in Frank Jackson's famous argument about Mary (Jackson (1982))--some challenge the argument by claiming Mary's changing her perspective by leaving the room does not give her any new information (see Lewis (1983)). Unfortunately, I do not have space to adequately address this issue here, although my own view is that Janet does not acquire any new information (I also agree with Lewis that Mary does not gain any information upon leaving the black and white room).

[22] Cf. Moran (2001).

question", she might conclude that she *ought* to believe that she has never been betrayed by her sibling, that her sibling is a nice enough person, or that she has no clue why she believes this horrible thing about her sibling. It may be slightly odd but not unintelligible to hear her say something like "I know I should believe that my sibling did not betray me, but I just can't". In ordinary cases, however, when a person judges that a particular belief is unreasonable she thereby extinguishes the belief. In this case, Janet has a sense of what a good reason to believe is, of what she ought to believe, but that assessment makes no difference to her psychological condition. Janet is more like a passive bystander or an observer of the belief that she was betrayed by her sibling; she happens to find out about it, but only after a great deal of effort.[23]

According to Moran, Janet admits no authority for her belief. Although we can see that Janet does lack authority, the missing authority looks to be connected with an ability or capacity to determine what she does believe on the basis of good, evidential, justifying reasons. Contrary to what Moran thinks, the missing authority does not seem to be epistemic.[24] As we saw, Janet could know perfectly well what her own beliefs about her sibling are, she could even be better placed to know than anyone else, but she would still lack a kind of authority that the rest of us normally have. Her case therefore suggests that there is an exclusively first-personal kind of authority but it is connected to our agential capacities rather than to epistemic ones.

III

The main difficulty with taking first-person authority to be a kind of epistemic authority is that it no longer seems to be a uniquely first-personal phenomenon. Epistemic accounts of our authority distort the *explandum* because they explain something that is not fundamentally tied to the first-person point of view. In this section, I shall develop this line of criticism in more detail by considering specific epistemic views.

Epistemic accounts of first-person authority have traditionally come in two main forms. The first I will call the "private objects model" and it is based on the idea that an individual's psychological states, or at least most of them, are fundamentally inaccessible to others. The idea is that the nature of psychological states in such that they are epistemically accessible only to one person, their owner. On this view, we can easily understand why each person will always be in a privileged

---

[23] It is true that Janet could causally force a change in her belief. She may take pills or hit her head repeatedly with a hammer. Although there is some agency involved here--Janet is doing something--she is merely hoping that what she does changes her beliefs. She cannot, in this manner, directly change her mind.

[24] Moran (2001), pg. 128.

epistemic position to know about her own. The second approach is proposed by philosophers who are unsatisfied with the private objects model. Contemporary philosophers prefer to explain first-person authority by appealing to some kind of privileged epistemic method or procedure that a person follows when she knows about her own psychological states. I will call this the "privileged method model" and most views fall into this category.

Let me begin with the private objects model. Some philosophers have already noted that this model is metaphysically hopeless and that it leads to skepticism about other minds. Although these difficulties are both real and insurmountable, I will not go into them. Rather, I want to show how the private objects model distorts the kind of authority we are trying to understand. According to the private objects model, metaphysical facts about our psychological states explain why each of us has access to only our own. Every person possesses epistemic authority because these are essentially private mental objects. But, the manner in which we come to learn about these allegedly private facts is fundamentally third-personal in form because it does not require that one adhere to either the *Davidson Principle* or the *Evans Principle*.[25]

If the nature of the psychological states fully explains why only I can be aware of my beliefs, I would need only to own them in order to enjoy epistemic authority with respect to them.[26] For example, consider Russell's early view that held psychological states are essentially "things with which I have acquaintance, things immediately known to me just as they are." (1912, pg. 47) If Russell were correct, the first-personal way of relating to those states would make no difference to my having authority. According to Russell's view, I should be directly acquainted with all of my psychological states and therefore have epistemic authority for any that I own. But, this is not the case. As we have seen, the kind of special authority we are considering depends on whether we related to our psychological states in an engaged first-personal way. The problem with the private objects model is that it explains

---

[25] David Armstrong (1968) thinks this is a significant advantage because it helps to demystify the mind. Thus, he wishes to describe self-perception as completely analogous to sense-perception. In this chapter, I am taking it for granted that our first-personal way of knowing is different from sense-perception, but perhaps the private objects model will look more appealing to someone who denies this.

[26] Moran has a similar objection in mind when he says that "nothing especially first-personal is captured by transferring the situation of a spectator from the outside to the inside, nor by construing the person as having any kind of especially good theoretical access to his own mind." Moran is not suggesting that someone could not, in fact, offer us some sort of explanation as to why a person's authority extends to only her own beliefs; rather, I think he is getting at the idea that any account along these lines leaves out why the ordinary first-personal way of relating to our beliefs matters for this authority.

something that I can have without being related to my psychological states in any distinctive way.

Since most philosophers now reject the private objects model, let's consider whether the privileged method model fairs any better. The idea motivating this approach is that first-person authority is based on some epistemically privileged procedure a person follows when she acquires knowledge of her own psychological states. Proponents of this approach rightly notice that there is a distinctive first-personal way of knowing one's own psychological states. Since this method is unique to the first-person, the model might appear to remain faithful to the first-personal character of our authority. But it also assumes that first-person authority has an epistemic sense. If we take that sense for granted, it can easily seem that an explanation of the special way a person knows about her own psychological states is also an explanation of her authority. But, crucially, this overlooks that we need not interpret first-person authority to have an epistemic sense. Because of this oversight, a proponent of the model takes for granted that first-person authority is a consequence of our having a distinctive way of knowing without characterizing either theses in terms that are especially first-personal. This leaves us with a picture that minimally recognizes the existence of a first-personal way of relating to our own psychological states but does not base our authority on any first-personal aspects of that relation. For this reason, I think the privileged method model ultimately fails to account for a fundamentally first-personal kind authority.

My diagnosis of what goes wrong with the privileged method model has been slightly programmatic, so it will be helpful to consider a couple of specific views that exemplify the model. Jane Heal offers what she calls a constitutive account of first-person authority. Her view has one core thesis: "the existence of a second-level belief about a first-level psychological state is itself what makes it true that the first-level state exists." (2001, pg. 4) Heal also asserts we cannot "just be mistaken" about our own minds, which shows that she takes the epistemic sense of authority for granted. Her purpose in her paper, as she states it, is to provide an account allowing her to combine the following three claims:

> (1) People's first-person present tense self-ascriptions of their psychological states are authoritative, while second and third person ascriptions are not authoritative.
>
> (2) Psychological predicates have the same meaning and role in first, second and third person uses.
>
> (3) Persons and their psychological states are among the public and effect-producing occupants of the universe. (2001, pg. 3)

Heal thinks that the three claims are paradoxical because she thinks it is puzzling how a person could be an authority for psychological states that are publicly observable. Part of her trouble here comes from assuming that "authoritative" as it occurs in (1) has an epistemic sense, in thinking she must explain, as Heal puts it, how we cannot "just be mistaken" about our own minds.[27] If psychological facts are publicly available, how could one person just *be* in a better position for knowing about them than everyone else? The only way to reconcile (1)-(3), according to Heal, is to adopt her constitutive thesis

Heal's primary concern is with making (1)-(3) coherent and she therefore overlooks a prior question. What makes (1) true? In what sense are self-ascriptions of psychological states authoritative? Let us suppose she is right that my second-level beliefs constitute my first-order psychological states. If I believe that I believe that it is raining in Berkeley, I make it true that I believe it is raining in Berkeley. Is this kind of constitutive relation sufficient for explaining my authority? If something about the ordinary first-personal way of relating to one's psychological states were necessary for the constitutive relations Heal has in mind, we would be heading in the direction of an explanation but that would require much more of an argument than the one Heal provides. As it stands, Heal's view cannot explain a kind of authority that depends on the first-personal way we relate to our own psychological states.

We can see this because Heal's account permits violations of the *Davidson Principle*. It is completely consistent with her view that I base all my psychological self-ascriptions entirely on behavioral evidence. Neither Heal's constitutive thesis nor the truth of (1)-(3) require that I form my second-level beliefs independently from an evidential considerations. For all Heal has said, I could form beliefs about my own psychological states because of what you tell me or because of the pronouncements of an oracle and have these second-level states constitute the embedded first-order psychological states. But there is no first-person authority in these cases. Heal gives us no reason to think her constitutive thesis would only be true if we form second-level beliefs in accord the *Davidson Principle*. She could say that second-level beliefs have this constitutive role only when they are so formed, but that kind of restriction is *ad hoc* .It also gets us no further in understanding why the first-personal way of relating to our own beliefs secures our authority.[28]

---

[27] Heal initially describes our authority as the phenomenon that, "what people say, in the first-person and present tense, about their own thoughts is treated as authoritative." But, a few lines later she glosses what she takes to be the same idea as the fact that "about one's own thoughts, however, one cannot, in this everyday manner, just be mistaken. In this way, Heal quickly slides into understanding first-person authority as our inability to be mistaken and thereby prematurely assumes the epistemic sense of authority.

[28] As I hope will be clear, once we understand more clearly what makes (1) true, its coherence with (2) and (3) will be obvious and not at all paradoxical.

I am going to consider one more example because I want to be clear that the kinds of problems I am raising are not peculiar to Heal's constitutive account but apply to any view that understands first-person authority to be a kind of epistemic authority. Alex Byrne attempts to explain our first-person authority in terms of a special method of following what he calls an epistemic rule. As Byrne describes it, an epistemic rule is a conditional of the following form:

R: If conditions C obtain, believe that p. (2005, pg. 23)

According to Byrne, a person follows R if she believes that *p* because she "recognizes" that conditions C obtain. For example, if conditions C are "it is raining", then I can follow an epistemic rule like R by believing that it is raining on the basis of seeing that it is.[29] Following epistemic rules is therefore not very difficult. What Byrne wishes to argue for is that there is a special, first-personal, epistemic rule, which he calls BEL:

BEL: If *p,* believe that you believe that *p*.

Suppose that it is raining and that I recognize this fact by looking outside the window. This amounts to fulfilling the antecedent of the BEL rule; so I ought to believe that I believe that it is raining. Byrne argues that BEL is a good rule to follow because it is self-verifying. He thinks that "recognizing that *p* is (inter alia) coming to *believe* that *p*." (2005, pg. 26) So, I cannot fail to believe that it is raining when I recognize that it is. I cannot even *try* to follow BEL without making the consequent of the rule true. As Byrne puts it, "because BEL is self-verifying, the truth of one's second order belief is guaranteed." (2005, pg. 27) The question, however, is whether this self-verifying rule formalizes a rule-following procedure exclusive to the first-person.

Following BEL may reliably generate true beliefs about one's own beliefs but not because BEL captures something distinctive about our first-personal way of engaging with our beliefs.[30] BEL can be true no matter how a person follows it. Suppose that every time it is raining, I hit myself in the head with a bat in order to

[29] I am assuming, hopefully harmlessly, that seeing is a way of recognizing.

[30] BEL generates true beliefs because it is self-verifying. In order to fulfill the antecedent of the conditional, I must, "recognize that P" which, Byrne tells us, comes to the same thing as forming a belief that P. So, I must *do* some believing in order to even attempt to follow the BEL rule. This somewhat deflates Byrne's point about the self-verifying character of BEL because other epistemic rules that require us to believe something in order to follow them are also self-verifying. Consider, this one: BEL': If P, believe that someone believes that P. BEL' would also generate true beliefs.

get myself to believe that I believe that it is raining. Seeing the rain is a way of recognizing that the conditions of the BEL antecedent hold, and hitting myself is a way that I could make the consequent true and follow the rule. But there would be no first-person authority in cases where I followed the rule in this way. Even if it is true that following BEL in a peculiarly first-personal way secures a measure of authority, the reason for this will not be because BEL is self-verifying. Because one can follow it in a variety of ways, BEL alone cannot account for an essentially first-personal kind of authority.[31]

The point I have been emphasizing throughout this discussion is that the phenomenon of first-person authority is essentially first-personal. The authority we are trying to understand is something we expect to be present wherever we encounter a person engaged with her beliefs. It seems that there is something special about this first-person perspective that in principle makes it authoritative and that is precisely what an epistemic approach fails to acknowledge. The most popular epistemic views ultimately describe a relation that is third-personal in form, a relation that a person could stand in to beliefs she is alienated from. If the phenomenon of first-person authority is something fundamentally connected to our nature as persons, interpreting that authority in an epistemic sense only leads us away from understanding the phenomenon.

IV

If what I have argued so far is true, there are difficulties with thinking of a person's first-personal relation to her own beliefs as one where she is in the best position to view or to offer an opinion on them. Perhaps instead it is better to approach what is special about our first-personal way of having beliefs in non-epistemic terms. Persons are usually thought of as responsible for their own beliefs; we assume they can make a substantial difference to what they believe. Normally, a person's own

[31] Near the conclusion of his paper Byrne seems to think that a "causal transition between mental states," can underwrite the ostensible rule-following behavior. At times, then, it seems that Byrne does not think a person consciously follows BEL at all but that a sub-personal causal mechanism is responsible for behavior that superficially accords with BEL. I say "superficial" here because, literally, there would be no rule-following going on at all. If an "appropriate causal mechanism" is responsible for generating higher-order beliefs, then it cannot strictly be said that the *person* is following any rules. If this is truly the direction that Byrne wants, then we are even further than from an explanation of first-person authority. There is absolutely nothing about a sub-personal "appropriate causal mechanism" that is essentially first-personal.

sense of what are good reasons for believing enable her to do something with respect to her beliefs, to form them or revise them or, in other words, determine what they are for those reasons. This is something that other persons simply cannot do, which makes it exclusive to the first-person point of view. Furthermore, as we saw with Janet's belief that her sister betrayed her, when a person relates to her own beliefs third-personally she seems to lack this ability to determine what she believes on the basis of reasons.

These and similar kinds of considerations motivate seeing first-person authority as a type of agential authority. Recall that a person has agential authority by being in the best position, and often the only position, to do a particular thing. I will argue that a person's capacity to change or maintain what her beliefs are directly on the basis of what she takes to be good reasons for them secures her agential authority for them. But only when I relate to my beliefs in an engaged first-personal way, am I a cognitive agent who is able to immediately affect their existence and character. As in Janet's case, when a person is detached or alienated from her own beliefs, her own judgments about what she has reason to believe do not determine what she does believe.

Unlike someone who is detached from her belief in a third-personal way, a person engaged first-personally with her belief has an attitude fundamentally directed at the truth.[32] Beliefs are and should be responsive to good reasons for believing, which are, because of the nature of belief, necessarily evidential reasons.[33] If a person has a belief that P and is presented with what she takes to be conclusive evidence that P is false, she will, unless she distances herself from the belief, immediately stop believing that P. Similarly, a person normally need only judge that there is good reason to believe that P and she will thereby immediate believe that P. This is not always the case. A person may, like Janet, have beliefs that do not respond to her judgments about what she has reason to believe. Nonetheless, when a person engages with her own beliefs in a first-personal way, the existence and character of those beliefs will depend on what she takes to be reasons for believing. We all take beliefs to be states responsive to truth in this way and when we attribute beliefs to others we presume that they will respond appropriately to the relevant reasons for believing.[34]

---

[32] Shah and Velleman (2005) make this point by saying beliefs are governed by a "standard of truth".

[33] I am assuming that an evidential reason is a reason in favor of the truth of the proposition one believes. For an argument to this effect see Shah (2006).

[34] There is a nice discussion of this aspect of belief in Chapter 4 of Stroud (2011). As Stroud rightly points out "we make sense of such a person's believing the things he does by finding that he comes to believe them on the basis of what he regards as reasons to believe it." (pg. 104) In that chapter, Stroud also draws on an analogy

The fact that beliefs are normally responsive to truth-conductive reasons helps explain the *Evans Principle*. When we relate to our beliefs in the first-personal way, we do not treat them as psychological facts independent from the truths in the world that they represent. We direct our attention to what Evan's calls "outward phenomena". When asked whether we believe that P, we attend to because facts in the world that we think ought to determine what we do believe. We look to considerations bearing on the truth of P because these are good reasons for believing. Although someone like Janet can also look to these "outward phenomena" when asked about her belief, because her belief that her sibling betrayed her is unresponsive to what she thinks she ought to believe, she cannot adhere to the *Evans Principle*. In this sort of case, if Janet were to try to answer a question about her belief in accord with the *Evans Principle*, if she were to try to answer by considering only the relevant "outward phenomena", she would give the wrong answer. She would say that she did not believe that her sibling betrayed her. However, a person who relates to her belief in a first-personal way constitutes what she does believe by making judgments about what she ought to believe. This is why such a person is justified in answering questions about her beliefs in the way Evans suggests.

It is important to understand that in order for me to be committed to the truth of my own beliefs I do not have to have actually formed them by deliberating on reasons. Many of a person's beliefs are causally produced without her deliberation or judgment. Nevertheless, a person is committed to the truth of her beliefs by being *disposed* to adjust them accordingly in the face of evidential or justificatory reasons. When questions about the reasons for my beliefs do arise, as perhaps when someone asks me a question about them, I may exercise my capacity to actively adjust or abandon my beliefs. But, if the question does not arise, I remain committed to the truth of my beliefs in the sense that my beliefs would be responsive to what I take to be good reasons--I would immediately change my belief that P in the face of compelling evidence that not-P.

In everyday life regardless of how my beliefs are actually formed, only I have a capacity to take responsibility for them by being committed to their truth in this way. My ordinary first-personal relation to my own beliefs is my being committed to the truth of what I believe. Another person can ascribe beliefs to me while being indifferent to whether or not they are true. She does not have to assess the reasons for or against what I believe in order to attribute them to me; she can accurately attribute beliefs to me on the basis of my behavior. But, in the first-personal case, I cannot be indifferent to the truth of my beliefs. My beliefs express my commitments to the truth and I cannot attribute them to myself on the basis of evidence without suspending that commitment. If, in order to self-ascribe the belief that P, I first consider something that does not bear directly on the truth of P, I acknowledge that

---

between believing and desiring and intending. This type of analogy will hold, I think, for an account of our first-person authority for those kinds of psychological states.

the existence of my belief depends on more than whether or not P is true. This is helps to explain the *Davidson Principle*. Since I must treat my beliefs as states whose existence depends only on the truths they represent, I cannot self-ascribe beliefs on the basis of evidence for their existence. Doing so would be a way of disengaging from them, of treating them as facts whose existence did not completely depend on what I take to be true.

The first-personal way of relating to our own beliefs is therefore essentially connected to our cognitive agency. I am the only person capable of engaging with my own beliefs in this way so that my sense of reasons for believing directly determines what it is that I do believe. Having this capacity requires that I adhere to both the *Davidson Principle* and the *Evans Principle*; it requires that I stand in the first-personal relation to my beliefs. Although a person like Janet shows that a person can sometimes have beliefs she relates to in a third-personal way, this cannot be the case for most of a person's beliefs. If the majority of Janet's beliefs were insensitive to her assessment of what she ought to believe, attributing beliefs to her would begin to lose intelligibility. We would have a difficult time attributing a type of psychological state that we take to be fundamentally responsive to reasons, while knowing it is unresponsive to reasons in Janet's psychological life.[35] Normally each of us is engaged in a first-personal way with what we believe. I believe this special way we relate to only our own beliefs explains why we have first-person authority. We can specify how this first-personal way of having beliefs secures our agential authority as follows:

First-Personal Agential Authority: A person X has first-person authority for a belief B iff X is the only person with a capacity to directly determine B on the basis of justifying reasons.

Only *my* sense of what counts as a good reason to believe something can affect my beliefs in this way, can cause me to believe or cease to believe that something is so. You might have a better grasp on what is, in fact, good evidence or be far better at assessing reasons for believing, but this, by itself, has no effect on what I believe. This capacity to directly determine beliefs on the basis of reasons is unique to the first-person and it is why we each have a special kind of authority, first-person authority.

As the example of Janet showed us, beliefs to which we are related in a more third-personal way are insensitive to a person's sense of justifying reasons. Janet cannot immediately determine what she believes on the basis of the relevant facts about her sister. Someone might object, however, that Janet does have this capacity and she merely fails to exercise it. If so and Janet still lacks authority for her belief that her sister betrayed her, my account of first-person authority is wrong.

[35] cf. Stroud (2011), Ch. 4.

According to this line of objection, Janet would fulfill the second half of the previous bi-conditional, she would be the only one with the capacity to determine her beliefs for reasons, but she would not have first-person authority. This objection is not trivial because if anyone is going to have this capacity, it is going to be Janet.

Part of my reason for going into the case of Janet was to bring out how the capacity to determine one's own beliefs for reasons depends on the way a person relates to those beliefs. It does not depend on mere ownership of beliefs. I think it is clear that Janet lacks this capacity with respect to her belief that her sister betrayed her. She tries to follow the *Evans Principle* by considering characteristics of her sister but she can only judge that they support the opposite of what she actually believes. Her judgments about good reasons for believing, which in a normal case constitute beliefs, do not affect what Janet actually believes. Janet thinks that she ought to believe that her sister did not betray her but cannot get herself to believe it. I do not think there is a point to insisting Janet has some hidden capacity to determine what she believes when she is trying as hard as she can to exercise that very capacity and failing. Janet is simply incapacitated.

Another reason to think that someone like Janet lacks this capacity when she relates to her belief in a third-personal way is that it explains why a person like Janet necessarily lacks first-person authority when she takes up a third-person perspective. Recall that on an epistemic interpretation, Jane does not have to lack authority when she takes a third-person stance toward her own beliefs. As we saw, she could acquire a kind of third-personal expertise that enables her to best assess evidence and thereby have a kind of epistemic authority despite violating the *Evans Principle* and the *Davidson Principle*. On any epistemic view, the connection between Janet's authority and her standing in a first-personal relation to her belief is obscure. It is not clear why taking up a third-person point of view should undermine Janet's authority. But, if Janet's capacity to determine her beliefs for reasons depends on her relating to them in a first-personal way, we can understand why she lacks authority for her beliefs when she relates to them in a third-personal way. By adopting the agential account over the epistemic account, we can begin to see these fundamental connections between our authority and the first-personal way of relating to our own beliefs.

V

Richard Moran, in his book *Authority and Estrangement,* has also appealed to agency to explain first-person authority. Although I think there is much to be learned from Moran's book, his view faces some difficulties. This is primarily because, despite his emphasis on agency, Moran takes first-person authority to be an epistemic

phenomenon.[36] Indeed Moran claims that a central task of his book is to explain why psychological self-ascriptions made in the ordinary first-personal way "enjoy a particular epistemic privilege not accorded to corresponding third-person judgments."[37] One problem for Moran is that agency alone cannot explain this. It cannot account for an epistemic privileged way of knowing about our own psychological states.

On Moran's view, if we are rational believers, subjects with beliefs that are responsive to the appropriate reasons, we are entitled to answer questions about our own beliefs without having to appeal to evidence. Instead, a person can answer questions about "whether he *believes* that P in the same way he would address himself to the question whether P itself. From the first-person point of view, the one question is treated as 'transparent' to the other."[38] When a person answers questions by following this "transparency" procedure, they issue something Moran calls an "avowal".[39] Saying that we should answer questions about our beliefs by "avowing" is a way of saying that we should adhere to both the *Evans Principle* and the *Davidson Principle*. Moran thinks we can do this because, as rational believers, we must assume that we can "make up our minds" about what we believe.[40] Although I agree that we

---

[36] He does not take it to be *purely* an epistemic phenomenon. Rather, Moran thinks that first-person authority has both what he calls an epistemic and an agential dimension. But, he also seems to think, mistakenly in my opinion, that the two dimensions are interdependent. See Moran (2001), pg. 92.

[37] Moran (2001), pg. 10.

[38] Moran (2004), pg. 457. See also Moran (2003), pg. 410. Because a question about one's beliefs is treated as "transparent" to a question about the world, Moran calls this the "Transparency Condition". I prefer to avoid the term "transparency" because to it conjures images of psychological states being self-intimating (see, for example, Bilgrami's (2006) use of the term).

[39] This is obviously a special sense of "avowal". Moran defines it as a statement of one's beliefs that is made in accord with this procedure. Moran (2001) pg. 101.

[40] Many times Moran writes that our rational agency entitles us to "assume" that we are able to answer questions about our own psychological states without an appeal to evidence. For instance, he claims that "a person is entitled to this *assumption* insofar as his answering the question proceeds from the *understanding* that his sense of the reasons in favor of P itself does determine what his belief about P is." (2004, pg. 466, my emphasis) The idea that we must "assume" something like this before we adhere to the *Davidson Principle* strikes me as incorrect. It makes it seem as if sophisticated aspects of our cognitive agency must be apprehended in order to legitimize our non-evidential self-ascriptions. I think it is better to say that we, as agents, are entitled to adhere to the *Davidson Principle* without having to assume anything at all. But perhaps this is all Moran intends by his talk of "assuming".

are entitled to answer questions just as Moran indicates, he also seems to think that this amounts to our "enjoying a particular epistemic privilege".

In part this is because Moran thinks first-person authority is connected to our way of accessing our beliefs. He writes that avowing is a "form of awareness"; it is a way of having special, first-personal, non-evidential access to ones own beliefs. Interestingly, Moran does not think that we have this kind of access to all our beliefs; we only have it "where it *is* a question about oneself that is being asked and answered."[41] But, when such a question is asked, answering by avowing, in Moran's special sense, is the way we are aware of our beliefs.

I think it is a mistake to think that a "form of awareness" or access can be fully explained by a procedure we follow to answer questions about our beliefs. It also seems wrong to me to say that avowing *is* a form of awareness. It rather seems that we must presuppose some first-personal form of awareness in order to answer questions by avowing. There would be no point to my trying to answer questions about whether I believe that P by considering facts that bear on the truth of P if I did not have some kind of first-person access to my cognitive actions. For it is not a passive kind of consideration that determines whether I believe that P but my judging whether or not these facts show that P is true or likely to be true. That is, I must consider facts to be reasons for believing P. My judgment that they are good reasons for believing P need not be explicit. As I have said, I may have a sense of the appropriate reasons for believing.[42] But it is my judgment about good reasons for believing that determines whether or not I believe that P, not P itself and not facts relevant to the truth of P that I do not understand as being so relevant. I must engage cognitively with the facts that bear on the truth of P in order to appreciate that they are reasons for me to believe that P.

But, this seems to require some distinct form of awareness or access to my own cognitive activity. It seems that I must be aware, for example, that I take Q to be a reason in favor of believing P. This is especially true if my consideration of Q is going to be relevant to answering questions about whether or not I believe that P. If I were not aware of what I judged to be a good reason for believing, no amount of attention paid to external facts could help me learn what I believe.[43] If, as Moran claims, part of our entitlement to avow rests on our capacity to "make up our

---

[41] Moran (2003), pg. 415.

[42] These judgments should not be thought of as a response to a question. Nor should they be thought of as overly intellectual or deliberative. In a very ordinary way, I can look outside and see that it is raining and also judge that it is raining. But I do not have to see it *before* I judge it. Afterward, I may also declare to you that it is raining, but my declaration should not be confused with my judgment.

[43] There is an analogy with *modus ponens*. I can "consider" both P and (P □ Q) but that alone will not determine that I believe Q. Rather, I must judge (implicitly) that these are conclusive reasons for believing Q.

minds", part also rests on our having awareness of what we are "making up" when we exercise this capacity.[44]A first-personal form of awareness is a necessary condition for answering questions by avowing and so avowing cannot be equivalent to that form of awareness.

For the moment, let's set this last worry aside and suppose that Moran can explain first-person access to our beliefs as he intends. Does this amount to our having epistemically *authoritative* or privileged access? The way we know about our own beliefs is clearly distinct; it is unlike our epistemic access to other contingent facts. But is there any reason to think it is epistemically privileged? Moran never attempts to explain why it is superior to the ways that other people know about our beliefs.[45] He rather seems to take its epistemic authority for granted.[46] But, if we do not simply assume that the way we know our own beliefs is privileged over other ways of knowing, is there anything about the features of agency Moran focuses on that would entail that it is?

Suppose, contra Moran, that our way of knowing our own beliefs is not epistemically privileged. I do not think this would undermine any of the central aspects of Moran's account. It would still be true that, when faced with questions about our beliefs, we are entitled to adhere to the *Evans Principle* and the *Davidson Principle;* in Moran's terms, we are entitled to answer these questions by avowing. Adhering to these two principles, seems to require having a special kind of first-personal, non-evidential access. Yet, by hypothesis, it is not epistemically privileged over other forms of access. That does not inhibit our answering the question whether we believe that P "in the same way" we would the question whether P. Contrary to what Moran seems to think, the features of cognitive agency responsible for our capacity to answer questions in this special manner do not require that each of us "enjoy a particular epistemic privilege" on our own psychological states. If first-person authority is derived only from these fundamental aspects of our cognitive agency, it is not an epistemic phenomenon. What Moran fails to realize is that,

---

[44] Peacocke (2008) makes a similar claim. He argues that having a capacity to engage in what he calls "mental actions" requires having a special kind of access to those actions.

[45] One reason it may be thought to be superior is because of the entitlement Moran argues is involved in coming to know one's own beliefs. But that would constitute an epistemic privilege only if there were no analogous entitlement for our way of knowing about another's beliefs. For an argument that there is such an entitlement, see Burge (1993).

[46] There are surely assumptions behind why he takes this point for granted. One of which may be that our access to other minds is inferior because it is mediated by perception. This assumption is shared by other philosophers (e.g., Bilgrami (2006)). However, it is a substantial assumption. It is also one that I believe to be false. I shall take up this issue directly in Chapter 5.

properly understood, features of cognitive agency can sufficiently explain how we are authorities in a purely agential sense.

Moran might think that having a distinctive from of awareness is equivalent to having a privileged way of knowing about one's own beliefs. He might think that if I have a special mode of accessing certain facts that is, in principle, unavailable to others, I thereby enjoy an epistemic advantage over them.[47] Despite the attractiveness of this line of thinking, the inference is fallacious. Having a distinct mode of awareness doesn't entail that it is epistemically privileged or a better way of knowing. There is a strong temptation among philosophers to think otherwise. I think this is because the inference *would* be valid if we made further assumptions about the nature of our first-personal way of knowing. Two assumptions in particular seem to be quite common. First, it is often assumed that the existence of psychological states entails first-personal knowledge of their existence; psychological states are though to be, as Ryle described them, "self-intimating". The second assumption is that our first-personal way of knowing is more reliable than any way of knowing requiring perception. Both of these, however, are substantial assumptions about our first-personal way of knowing. Either one, if true, could help explain a person's "particular epistemic privilege" over others but, without one of these assumptions, Moran's account cannot. I believe both assumptions are mistaken and will discuss them at some length in Chapters 4 and 5 of this dissertation. At this point, it is sufficient to notice that Moran cannot sufficiently explain a person's "particular epistemic privilege" over others.

Some philosophers have voiced a very different objection to Moran's view. They argue that we cannot always "make up our minds" because many beliefs are not really deliberatively up to us. Shoemaker puts this objection to Moran as follows:

> I know that I am wearing pants, so I believe that I am wearing pants. I know and believe that I believe that I am wearing pants….But it is hard to think of circumstances, other than those of a dream, in which it could be a question for me whether I believe this. I would also have a hard time saying what reasons I have for believing it. And I cannot think of any good sense in which it is 'up to me' whether I believe it.[48]

---

[47] One place where this inference is explicit is in Boyle (2009), who writes: "self-ascriptions of these kinds of mental states are not normally liable to the same kinds of *error* that afflict ascriptions of such states to other people," which Boyle writes is "a manifestation of their *authority*." But if my self-ascriptions are not liable to the "same kinds of *error*" they may nevertheless be subject to different kinds of error. One should not infer that a way of knowing is epistemically privileged just because it is unique. Cf. Peacocke (2008).

[48] Shoemaker (2003), pg. 396.

It seems to me that there are two separate issues in what Shoemaker says here. Shoemaker highlights the familiar fact that we cannot adopt or discard our beliefs by voluntary acts--we cannot choose to believe anything whatsoever. If that were the type of agency required for first-person authority, an agential account would be hopeless. We can easily avoid the worry because, as we have seen, beliefs are fundamentally constrained by evidence, reasons, and justification.[49] Yet Shoemaker also objects that much of the time we cannot even make up our minds in this sense. Many beliefs, like the one that I am wearing pants, are ones I simply find myself with.[50] I believe that I am wearing pants, that the earth exists, and many other things besides that I never deliberated about and probably never will.

Shoemaker's criticism presents a challenge to thinking of first-person authority in terms of agency. Many beliefs, as Shoemaker suggests, are not formed through anything like a deliberative process and could not come up for serious reconsideration or adjustment, even once we are aware of them. When it comes to a belief like the one about the earth exists, my mind is simply made up and it will fortunately stay that way no matter how much attention I pay to it. It is difficult to imagine what a rational reconsideration of this belief could even look like. Nevertheless, these are beliefs that I seem to relate to in a first-personal way, I am not detached or disengaged from them like Janet is from her belief that her sister betrayed her. I have argued that first-person authority is part of this ordinary first-personal way of believing but, how can I have first-person authority in an agential sense for beliefs I could not seriously reflect upon? How could these beliefs be determined by my sense reasons?

Shoemaker's objection brings out why we should set aside Moran's language that suggests a person needs to deliberate or be asked a question about her beliefs in order to have first-person authority. Determination of beliefs through deliberation would be a special cognitive activity but then we would have authority for only a few beliefs. First-person authority intuitively does not seem to require much reflection. It also does not seem right to think that another person must first ask me a question or that I must reflectively consider beliefs before I can possess authority for them. Fortunately the agential account that I have presented does not require deliberation. As I emphasized, we have a special capacity for agency whether or not we activate it through deliberation or reflection on what we believe.[51] My belief that I am wearing pants would be extinguished were I to see that I was not clothed. This is how my beliefs are supported by reasons even when they are not the objects of direct reflection. If the reasons never change, I would obviously not display my sensitivity to reasons for believing but I have this capacity nonetheless provided that my beliefs would respond appropriately in the relevant circumstances.

---

[49] See Moran's reply to Shoemaker (2003) and also Moran (2004).

[50] See Heal (2004) for the same criticism.

[51] See Raz (1997) for a helpful discussion of agency without direct action.

Sensitivity to reasons for believing is likely to come in degrees, to be something that has its own characteristic failures, good days and bad days. But, where we lose sight of it altogether is when a person abrogates her capacity for agency by adopting a third-personal point of view on her own beliefs. For any given belief that a person has, so long as her own sense of good reasons for believing determines whether she maintains, reforms, or abandons it, it is one for which she has authority. Only when it comes to one's own self are we able to determine beliefs in this way, directly on the basis of reasons for what we ought to believe. And, since I am the only one who can do these things, I have a special kind of first-person authority.

CHAPTER 3
ON DEFERRING

In our everyday conversations with other people, we usually defer to what a person says about her own psychological states. If someone comes up and tells us, for instance, that she believes that the Giants are going to win the World Series or that El Cerrito is a nice place to live, we presumptively treat these claims as true. This presumption of truth may be overridden by familiar considerations ranging from psychological pathologies to bald-face lies. In that sense, what someone says about her own psychological states is not guaranteed to be true. Nevertheless, when a speaker is sincere and engaged with her thoughts in a normal first-personal way, her assertions about them enjoy a unique standing in conversations, a standing that appears to warrant deferring to her psychological self-ascriptions.

Deferring to another person's psychological self-ascriptions is a way of conferring a special *epistemic* status on them. Out of all the assertions a person could make about contingent facts, only those about her own psychological life are immediately treated by us as true statements. This practice of deference would clearly be rational if we were picking up on the epistemic status of the person's assertions, if they provided us with the best possible evidence as to the existence and character of her own psychological states. Why else would we be justified in deferring to what others tell us about their own psychological states if not because they know what is going on better than anyone else does? If my psychological self-ascriptions were not connected to some sort of epistemic privilege, it is unclear why they should be accorded such a special epistemic status in conversations.

This practice of deferring is easily recognizable and it is perhaps most responsible for suggesting to philosophers the presence of some kind of first-person authority.[52] We seem to defer to what others say about their psychological states because in this domain they have special authority. Traditionally, this authority has been understood as an epistemic kind of authority, consisting in the fact that each person is in a privileged position to know about her own psychological states. But, in the previous chapter, I argued that the standard view mischaracterizes first-person authority by describing something that is not essential to the first-person point of view. I argued that instead we should understand it as a kind of agential authority. One thing that is puzzling about my proposal, however, is that if a person is not in the best position to know about her own psychological states it is not clear why would we should presume that what she says about them is true. Without first-person *epistemic* authority, the entire practice of deferring to what someone says about her psychological states appears to be unwarranted.

---

[52] For example, Davidson (1984), Heal (2001), Smith (1998) and Wright (1998) all point out this phenomenon of deference as indicating some kind of first-person authority.

In this chapter, I shall argue that deferring to a person's psychological self-ascriptions is justified by the agential authority each of us ordinarily possesses. Although deferring is a way of conferring a special epistemic status on a class of assertions, it need not be a response to the epistemic status of those assertions. In many contexts, deferring is a rational response to the epistemic properties of a speaker's assertion. This is true, for example, when we defer to a medical expert's assertions about a disease or neuroscientist's reports about the properties of our brains. In these cases, the epistemic status of the assertion is the primary, and usually the only, factor that could justify deference. But, when a person who is relating to her own psychological states in a first-personal way makes an assertion about them, she seems to be doing something other than reporting psychological facts.[53] Further reflection on the practice of deferring to psychological self-ascriptions reveals additional subtle ways in which it is distinct from the kind of deference we show epistemic experts. For this reason, it is plausible that deferring to an individual's psychological self-ascriptions is justified by non-epistemic properties of her assertions. In this chapter, I shall argue that when a person says something about her own psychological states, she presents herself as a cognitive agent with unique capacities to determine what those psychological states are. She does this by expressing, in what she says, her agential authority to constitute the very psychological states she is talking about. Deferring, I believe, is a rational response to this expressive property of psychological self-ascriptions, to its being an expression of one's agential authority.

[53] It might be objected that a speaker is, for this reason, *not* making an assertion. This is, for instance, what I think Bar-On (2004) intends when she says that psychological self-ascriptions are expressions "whose point is not to offer a descriptive report, make an assertion, or provide someone with information about the speaker's present thoughts, feelings, emotions or attitudes." (pg. 260) If I thought that an assertion was only a kind of report of information, then I would see the point of making this kind of distinction (for various ways of thinking about the nature of an assertion see MacFarlane (2011)). However, it seems to me that Williams is right when he says "if a speaker comes out with a declarative sentence not as part of a larger sentence (as one might say, by itself) and there are no special circumstances, then he is taken to have asserted what is meant by that sentence." (2002, pg. 74) Openness in speech just seems to issue in an assertion. Perhaps this is a weak enough notion of assertion to be agreeable to Bar-On. It is at any rate, a fairly meager notion of assertion that I intend throughout this chapter when I use the term.

I

Deferring to what someone says is not unique to psychological self-ascriptions. In many situations, it is appropriate to defer to what a person sincerely says about some subject matter.[54]There are many cases of deferring to epistemic experts, to individuals who are, in fact, especially well placed to know certain facts or to learn about them in some way. To take an example, my mechanic often makes assertions about the engine of my poor performing car. When he does, I presume that what he says is true. If he says, "the timing belt is broken," it is reasonable for me to believe that the timing belt is broken. My belief is rational in this case because my mechanic is an epistemic expert on the car's engine. He both knows more and is in a better position to learn about this particular domain of facts. These are good reasons to trust what he says. In deferring to what he says about the timing belt, I am therefore responding to the epistemic properties of his assertion, to the fact that it is made by someone with relevant epistemic expertise. Because the epistemic properties of the assertion justify my taking it to be true, this is a *purely* epistemic kind of deference. It is a way of conferring a special epistemic status on an assertion because of its antecedent epistemic properties.

The kind of deference we practice toward psychological self-ascriptions does not seem to be of this pure epistemic kind. There are fundamental and easily recognizable differences between the two. Consider again purely epistemic deference. After dropping off my car, I defer to my mechanic's assertion and believe that the timing belt is broken. It seems clear that the epistemic context makes a significant difference to my deferring. The epistemic context partially determines whether or not my deference is warranted. From my point of view, I defer to my mechanic's assertion that the timing belt is broken because I know that he is a good mechanic. This is why I took my car to him to begin with; I wanted a suitable expert to assess my car's engine. If the epistemic context were different, I would not defer to his assertion.

Suppose that I were not in my mechanic's garage after dropping my car off for inspection but instead my car suddenly stops working in the parking lot of a nearby grocery store. As I stand there staring at the engine, a stranger, who is in fact my mechanic, dressed in a business suit walks over and looks at my engine. After a few moments, he says, "Your timing belt is broken". Let's suppose in this example that I do not know this person as my mechanic. If this sort of situation were to

---

[54] In any discussion of deference, we must, for obvious reasons, assume the speaker is sincere. In any context, if someone lies or is otherwise misleading, she loses her entitlement to deference. This is not to say we might not in fact continue to defer to what she says, only that we are wrong to do so. In the remainder of this chapter, I will therefore assume a speaker's assertions are sincere, which, I agree with Williams (2002), seems to be a normal feature of all assertions.

occur, I do not think I would defer to his assertion. Perhaps this is because I have encountered too many people pretending to have knowledge they do not actually possess or perhaps it is because of my philosophy background. But I think I would be skeptical. More importantly, it does not seem reasonable for me to defer to a speaker's assertions about my car engine without knowing anything about him or about the epistemic parameters surrounding his assertion. What reason could I have for trusting this person's claim that my timing belt is broken? To me, it does not seem that there would be any.

It might be said that I do have a reason to trust the stranger's assertion, it is the same reason that I have when I am in the garage because the stranger *is* my mechanic.[55] He has the same degree of epistemic expertise whether I know him or not. So, even if I do not in fact defer his assertions, *qua* stranger, I ought to. But, this is just another way of illustrating how the epistemic context affects the practice of purely epistemic deference. It would be because the assertion about the timing belt is made by a good mechanic who is able to base his assertion on an accurate assessment of relevant evidence that I ought to defer to it. The epistemic context of the assertion determines whether or not deference is appropriate. This is a basic feature of purely epistemic deference.

But what happens if the person who approaches me in the parking lot talks instead about his own psychological states? What if he were to say, for example, "I believe that El Cerrito is a nice place to live"? It seems to me that I would normally presume that what he said was true, that he did, in fact, believe that El Cerrito is a nice place to live. Assuming that he is sincere, it seems that I would defer to what he said about his belief without any epistemic considerations coming into play. It would not matter that the person was a stranger and it does not seem that I would need to know anything about his past reliability or level of epistemic expertise. The only thing that does matter for whether or not my deference is justified is that he appears to be a normally functioning person who relates to his psychological states in the ordinary first-personal way. The epistemic context surrounding his assertion does not seem to be in any way relevant.

Another asymmetry between the kind of deference we practice toward psychological self-ascriptions and the purely epistemic kind of deference can be seen in the way that an appeal to evidential support affects the two. The assertions of an epistemic expert can reasonably be supported by a speaker's direct appeal to evidence. Suppose that when I go to pick up my car I am in an unusually suspicious mood. My mechanic may say that my timing belt is broken, but I could doubt him. If I were to express these doubts, my mechanic could appeal to the evidence upon

[55] Even if this is the true, one may, for familiar reasons, wish to maintain that my not deferring is not irrational.

which he based his assertion.[56] He could appeal to evidential considerations to epistemically support what he had said. My initial doubts about his assertion would be trumped by his direct appeal to evidential support. When an epistemic expert makes a direct appeal to evidence, it strengthens the credentials of his assertion and thereby further justifies deferring to it.

Things are exactly the opposite in the case of deferring to psychological self-ascriptions. As we have already seen in the previous chapter, not only do we normally not require psychological self-ascriptions to be based upon evidence, it would not even really make sense to press someone for her evidence. This is because our ordinary way of self-ascribing psychological states is fundamentally characterized by a lack of evidential ground.[57] This is also partly why contextual epistemic parameters are not relevant to deference in these cases. But, although epistemic factors do not normally enter into my psychological self-ascriptions, they sometimes do. This can happen when I shift to a more third-personal perspective on my psychological states. Shifting into this perspective on my own psychology is like taking up a point of view on my psychological life similar to the one I must take on another person's. In thinking about what I believe, I could base my self-ascriptions directly on evidence. I could choose to base them, for example, on the diagnosis of my therapist or on my own observations of my behavior. But, when a person does base her psychological self-ascription on evidence, she seems to lose her entitlement to deference. If I tell you that I believe that El Cerrito is a nice place to live because I noticed that I own an El Cerrito t-shirt or because my therapist told me, I would naturally sound strange. Once you are told how I become aware of my belief, you will normally stop deferring to my self-ascriptions, doing so no longer seems appropriate. Thus, unlike cases of purely epistemic deference, a direct appeal to evidence in support of a psychological self-ascription actually seems to undermine a speaker's entitlement to deference.[58]

---

[56] Here I am slightly altering an example found in Heil (1992). Heil remarks that a person making a psychological self-ascription often is not able to appeal to supporting evidence. But, as I will suggest, there is more to this asymmetry of evidential support than what Heil indicates.

[57] This is, of course, the *Davidson Principle* from Chapter 2.

[58] This does not mean that a person basing her psychological self-ascriptions on evidence could not *acquire* an entitlement to a purely epistemic kind of deference. Recall the example of Janet from Chapter 2. She was imagined to become a kind of expert in psychoanalysis who could thereby make excellent self-ascriptions on the basis of evidence. If this were to happen, she would be entitled to *purely* epistemic deference. But it is important to see that this would amount to a change in the way we treat Janet's assertions. The initial direct appeal to evidence would undercut Janet's entitlement to deference precisely because it is not of the purely epistemic kind.

These differences bring out how one might be liable to misunderstand the claim that deferring to a person's psychological self-ascriptions is immediate. In noticing that we defer immediately to a person's psychological self-ascriptions, we are not merely noting a feature about our own psychological processes. Immediate deference should not be confused with instantaneous deference. The immediacy of our deference to psychological self-ascriptions should be understood as epistemic immediacy. One thing the previous considerations illustrate is that purely epistemic deference is not immediate in this sense. It may be true that my deference to my mechanic is psychologically immediate. Having known him for some time, I might, in a purely temporal sense, immediately believe his assertions about my car's engine. But, this kind of deference depends on, and is therefore in a different sense mediated by, epistemic factors. Even if I do not reflectively consider it, the epistemic context of my mechanic's assertion is always in the background, determining whether or not I ought to grant it a special epistemic status. In this way, purely epistemic deference is epistemically mediated even in cases where it is psychologically immediate.[59] By contrast, because first-personal psychological self-ascriptions are, in principle, not based on evidential considerations, epistemic parameters do not affect whether our deference to them is appropriate--in this way it is epimsteically immediate.

Reflecting on these asymmetries between kinds of deference suggests that when I relate to my psychological states in the ordinary first-personal way you do not respond to the epistemic properties of my self-ascriptions. If the reason we deferred to self-ascriptions was because of their epistemic features, this kind of deference would look more like the *purely* epistemic kind we practice with experts. It would seem to be warranted, at least partially, on the basis of contextual epistemic parameters. A careful examination of the phenomenon, however, seems to show that when we defer to psychological self-ascriptions we are engaged in a distinct kind of practice uniquely tied to the first-person.

One might reasonably wonder what the relationship between these two kinds of deference is. What would happen, for instance, if some other person were in a better epistemic position to know about my psychological states? Would people be justified in deferring to her assertions over my own? Deferring to this other person would be of a purely epistemic kind. So, the question is whether certain epistemic contexts might overrule the entitlement to deference that a person's own psychological self-ascriptions usually have. If so, then even if our practice is not a response to the epistemic properties of a person's self-ascription, certain epistemic parameters would set a necessary condition on the justification of the practice. This possibility presents an interesting issue that I will have to return to in the final section of this chapter. Before I can address it, however, we need a better

[59] The difference between these two kinds of immediacy is clearly noted in Fricker (1998).

understanding of what we are responding to when we defer to psychological self-ascriptions.

II

On the basis of the distinctions illustrated in the previous section, it seems that a fully epistemic approach to understanding why we defer to psychological self-ascriptions cannot be right. We do not seem to be responding to the epistemic properties of what someone says about her own psychological life when we presume that what they say is true. Instead, it would seem that we are responding to some non-epistemic properties of her assertion. But what other features could justify our deferring? What properties of the assertion are relevant? If this is a kind of deference that is not a purely epistemic practice, its justification can seem to be quite obscure.

One suggestion is that we defer to psychological self-ascriptions because of their expressive properties. Wittgenstein once wrote that the "the expression 'I believe that this is the case' is used like the assertion 'This is the case'" (1958). Because of this passage and others like it, readers sometimes read him as saying that assertions that superficially appear to report a person's psychological state are really nothing but expressions of those underlying states.[60] Although it may look like I am reporting my beliefs when I say "I believe that the Giants will win the Series," a proper analysis of my speech act is that I am only expressing the underlying belief <that the Giants will win the Series>. An expressivist approach to understanding deference will maintain that when a person self-ascribes a psychological state she is expressing the state, rather than issuing a report about the state. Moreover, she, as the owner of the underlying psychological state, is the only person able to directly express it. These features of her assertion are thought by some to explain why we justifiably defer to what the speaker says. If we pursue this line of thinking, a fair amount will hinge on what it means exactly for an utterance to be an *expression* of a psychological state.

The most naïve version of expressivism holds that all psychological self-ascriptions are verbal outbursts like moans, groans, and whimpers. Moaning, for instance, is a way for a person to express an underlying condition of pain.When someone moans, we understand that she is in pain but not because she has said that she is in pain. A moan is not a report of her pain. It does not even have truth-conditions. The naïve expressivist thinks of psychological self-ascriptions in the same way--as speech acts that, like a moan, lack truth-conditions.

---

[60] I personally think that this is not Wittgenstein's own view. But I do not wish to argue for that point. For a thoughtful interpretation of Wittgenstein on this point, I recommend Finkelstein (2003).

There are obvious shortcomings with the naïve view. Unlike moans and groans, there is plenty of semantic evidence indicating that psychological self-ascriptions do function like normal assertions in various ways. One example is that whenever I say "I believe that the Giants will win the Series," it seems to be synonymous with your assertion of "Matt believes that the Giants will win the Series." If that is true, then what I say cannot be a mere expression, like a moan, but must be an assertion of a proposition with truth-conditions, the same truth-conditions as your assertion. The naïve expressivist cannot make sense of this fact nor can she help us very much with other kinds of familiar semantic features of self-ascriptions.[61] For my purposes here, I am going to quickly set aside the naïve view without much argument because I think it obviously fails to accurately capture our linguistic practices. The naïve view, at any rate, has very few, if any, adherents. What is crucial to recognize from its brief consideration is that there is nothing about the notion of a person expressing a psychological state that entails that what she says cannot also be an assertion. When David Finkelstein develops his version of expressivism, he correctly points out that "to maintain that some utterance functions as an expression is not--or, anyway, need not be--to deny that it is an assertion." (2003, pg. 99) With Finkelstein's point in mind, we should turn to a more sophisticated style of expressivism. It may very well be that one way to express a psychological state is by making an assertion with truth-conditions which self-ascribes the same state.

The core of any expressivist thesis is a negative point, shared with the naïve view, about the type of action a speaker performs when she self-ascribes a psychological state. Dorit Bar-On writes:

> The point of the subject's use of words is not to offer a descriptive report of her state, or to provide evidence for its presence, to inform someone about it. The subject's act of self-ascription may have no other point than to vent her frustration, shout for joy, give voice to her fear, air her idea, articulate her thought, let out her anger, and so on. (2004, pg. 243)

This description of an "act of self-ascription" is intended to close off the possibility of an epistemic gap between what a person says and the psychological state she is talking about, the kind of gap wherein errors could arise. When someone performs an expressive act rather than reporting on the nature of her own psychology, she is simply venting or otherwise expressing her underlying psychological condition. As

---

[61] The naïve view is adequately criticized by Wright (1998), who brings up familiar arguments about embedding psychological self-ascriptions in subordinate clauses of conditionals. It is also effectively criticized by both Finkelstein (2003) and Bar-On (2004) in the context of their presentations of more sophisticated forms of expressivism.

Bar-On often puts it, she is speaking "from her state". There is, therefore, no point in searching for an epistemic method or procedure lying behind her self-ascriptions that could occasionally be mistaken.[62] For the sake of argument, let's grant that when a person utters a psychological self-ascription her speech act is a way of expressing one of her psychological states; it is an expressive act. How far will this get us in understanding deference?

Consider the point of view of listeners. When someone groans or moans, we do not ask her for reasons for her speech act. These are expressive outbursts and they cannot reasonably be subjected to epistemic criticism or evaluation. If psychological self-ascriptions are of the same kind, if they are also expressive acts, they will also be immune from epistemic assessment. This is a point that Bar-On repeatedly stresses:

> To the extent that we regard a subject as simply giving voice to the condition she self-ascribes, rather than, say, providing an evidence- or recognition-based report on her own self-findings, it should indeed seem inappropriate to ask after the reasons she has for the different aspects of the self-ascription she produces when avowing. To do so would be to betray a misunderstanding of the character of her performance.[63]

On Bar-On's view, a person's psychological self-ascriptions are immune from the epistemic assessment of listeners because they are taken to be expressive acts, like moans and groans. But being immune from epistemic assessment is not the same as being entitled to deference. It is true that practicing deference is a way of treating someone's assertions as immune from epistemic evaluation, but it is wrong to think that any kind of immunity is equivalent to deference. The phenomenon of deference consists in listeners conferring a privileged epistemic status on a class of utterances; when we defer we treat a psychological self-ascription as *true* or as something that ought to be believed. This is not the same as refraining from epistemically criticizing them. More importantly, it is *not* how we treat moans, groans, or whimpers. These kinds of expressive outbursts are indeed outside the realm of epistemic evaluation but they are not thereby being granted a privileged epistemic status. Cries of "ouch"

---

[62] Someone might be tempted to object that one's psychological self-ascriptions are based on an investigation of her psychological condition. But this would, I think, significantly mischaracterize the phenomenon because it would amount to denying the *Davidson Principle*. Even philosophers like David Armstrong (1968) who think of introspection like a kind of inner scanning do not seem to think that there is an epistemic gap between the results of the inner scanning process and the corresponding self-ascriptions.

[63] (2004, pg. 263). Also, "You are *speaking* from a present condition, instead of giving it some non-verbal expression." (pg. 300)

and moans cannot be presumed to be true and they do not take up a special place in the subsequent reasoning of anyone who hears them.

So, the expressivist must move beyond gesturing at the expressive character that psychological self-ascriptions might share with groans and provide an account of how speakers are justified in taking what someone says about her own psychological states to be *true*. More sophisticated expressivists have taken up this challenge:

> We propose to locate the basis of the presumed truth of all avowals, not just avowals proper, in the expressive aspect of avowing. In avowing, subjects use truth-apt self-ascriptive sentences that are presumed to state truths. Our proposal is that *the asymmetric presumption of truth governing avowals applies to those self-ascriptions that are seen as expressive performances. It is grounded in the presumption that subjects' avowals are transparent to their mental conditions.*[64] (Bar-On and Long(2001), p.p. 328-329)

This passage needs careful consideration. Suppose that when I say "I believe that it is raining", I am, *qua* speaker, expressing my belief about the rain. Bar-On and Long claim that all such expressive acts as "transparent" to their underlying mental condition. So, for example, when I express my pain by screaming "ouch", what I say, "ouch", transparently reveals or shows my pain. For this reason, when a person hears my cry of "ouch", she can thereby perceive my pain. The pain is literally perceivable in what I say. If I am performing an expressive speech act when I self-ascribe a psychological state, it too can literally be perceived in what I say. [65] So, on this view, my expressive utterance of "I believe it is raining" transparently shows listeners its underlying condition, which in this case just is my belief that it is raining. Since listeners can see my belief in my speech act, it would seem that they have an excellent reason to presume that my self-ascription is true. Because I am expressing my psychological state, they can perceive the truth-conditions of what I say.

If this is correct, two aspects of a psychological self-ascription explain why listeners are justified in deferring to it. My speech act must be an expressive performance; it must be my way of expressing an underlying psychological state. This alone is not sufficient because the expression of a psychological state could happen in a variety of ways. I could, it seems, express a belief without also semantically representing it in the content of what I say. For example, one way to express my belief that one of my students is not well-suited for graduate school is by saying "I believe that he really enjoys philosophy". Or, I could express my dislike for

---

[64] As Bar-On and Long use the term, "avowal" means a psychological self-ascription that is an expressive act, as opposed to one that is a "mere report". The proposal in this passage is also the one endorsed by Bar-On in her (2004).

[65] Another person who holds that another person's psychological states can be directly perceived in what they say is John McDowell (1982).

a certain philosophical account by saying "I do not like Platonism".[66] And, it seems incredible to deny that an excellent way to express my belief that it is raining is simply by saying "it is raining." For Bar-On and Long's explanation to succeed, it is not enough for my underlying psychological state to be expressively vented by my speech act. It must also be the case that the underlying psychological state is semantically represented in the content of my assertion.

It is important to notice how two distinct senses of "expression" are being used in this expressivist proposal. First, a person's self-ascription, *qua* speech act, is characterized as a way of expressing her underlying psychological state. But we can also understand a self-ascription as the *product* of a person's speech act, as the thing uttered by the speaker. This is a proposition with truth-conditions that can be said to *semantically* express a person's psychological state; but this is a different sense of expression. In order for the expressivist to explain the phenomenon of deference, it seems that one underlying psychological state must be expressed in both of these senses. That is, the proposition that I assert must semantically express the very same psychological state that my act of self-ascribing expresses in the non-semantic sense. The expressivist requires the following principle to be true:

Expressive Symmetry: When P is a self-ascription of a psychological state, a person's uttering P is a way of expressing her psychological state *x* only if P semantically represents her as having *x*.

For example, saying "I believe that P" is a speech act that expresses someone's belief *x*, only if *x* is the belief that P. If *Expressive Symmetry* were not true, it would be possible to recognize a self-ascription as an expressive act or performance without presuming what is said is true. This would be the case whenever the psychological state a listener perceived in the *act* of self-ascription was not semantically represented in the content of the product of that act.

Bar-On seems to think *Expressive Symmetry* is true. She claims that "we take it that she *is* in the relevant condition--the condition that is semantically referred to by the self-ascription, which is *the very condition that would render the self-ascription true*." (2004, p.p. 316-317) We treat self-ascriptions this way because "self-ascriptive verbal expressions *wear the conditions they are supposed to express on their linguistic sleeve*." (2004, pg. 315) According to Bar-On, for any act of self-ascription, the function of expressing or venting an underlying psychological state is played by the semantic content of what a speaker says. The function of saying "I believe that P" is to expresses one's belief that P. Thus, an act of self-ascribing is an expressive act, and regarded as one,

[66] There is a wonderful story I have heard about Gilbert Ryle. Ryle is asked about his views on some specific theory and Ryle replies "Don't much care for 'isms'. If the story is true, it seems to me that Ryle had no problem expressing his underlying psychological state.

on the condition that the spoken proposition semantically represents the speaker as having the psychological state she is expressing. It would also presumably not been seen as an expressive act if it did not "wear the conditions" it expressed on its "linguistic sleeve". This is the idea *Expressive Symmetry* is meant to capture.

The difficulty facing the expressivist is that it does not seem that *Expressive Symmetry* is true. Rather, it seems that a speaker can sufficiently express a variety of psychological states without having to semantically represent being in that state. Moreover, it also seems that an utterance that does semantically represent a person being in a particular state can non-semantically express a many different psychological states. I have already mentioned two examples of expressing an underlying belief or desire where a different psychological attitude is represented in the content of what is said. There does not seem to be a very strong restriction on the kinds of things a person can say in order to express her underlying psychological states. In the right context, with the right audience, it seems to me that someone may utter nearly anything as a way of expressing her underlying psychological states. It may be true that saying "I believe that Peter enjoys philosophy" is sometimes my way of expressing my belief about Peter's enjoyment, but at other times it could be my way of expressing my belief that he will struggle in graduate school. What psychological state a token self-ascription expresses depends on the context and what the speaker intends to express.

One might reply that we do not choose to express our pains by saying "ouch". Our pains simply cause expressive outbursts of "ouch", which is why "ouch" is a characteristic expression of pain. Similarly, it might be thought that self-ascriptions are causally produced by their underlying psychological states and this is why they are characteristic expressions of those states. This way of construing the expressivist thesis may be able to disregard the kinds of contextual parameters that might bear on what specific expressive role a particular utterance plays. The problem, however, is that on the causal understanding of expressivism we have even less reason to believe *Expressive Symmetry*. Causal relations can hold between any two things; so any psychological state could cause any sort of natural expression. For the causal account to explain our practice of deference, it would have to be the case that the utterance of a self-ascription is only caused by the psychological state it semantically refers to. But, especially given the prevalence of false self-ascriptions, this does not seem to be the case.

At this point, one might hope for a more elaborate explanation of *Expressive Symmetry* in evolutionary terms. We could, along these lines, hope to explain *Expressive Symmetry* by appealing to more primitive utterances that were used to express psychological states. In order to make this kind of explanation more plausible, we could relax *Expressive Symmetry* to a claim that held for the most part. We could then suggest that psychological self-ascriptions are presumed to be true because, in fact, they usually express the psychological state to which they directly refer. To support this suggestion, we would explain the existence of this regularity

with an evolutionary account of how, over time, self-ascriptions came to take over the functional role played by more natural expressions of psychological states.[67] I doubt that this kind of regularity exists. But, one who wished to pursue this evolutionary approach would, I believe, have to overcome an additional obstacle. This is the fact that the most natural way of expressing one's belief that, for instance, it is raining is by saying "it is raining". In fact, the self-ascription of the belief in speech strikes me as a far less natural way to express it. [68] It would therefore be remarkable, from an evolutionary point of view, that humans developed two distinct ways to verbally express the same underlying psychological state.

Expressivist accounts of deference also face a second serious difficulty. Like fully epistemic explanations of deference, expressivists cannot truly respect the essentially first-personal character of the special kind of deference we practice toward psychological self-ascriptions. Earlier, we saw that when a psychological self-ascription is based on third-personal or public evidence it does not seem to warrant deference. It therefore seems that something about standing in a first-personal relation to one's psychological states licenses the kind of deference we are trying to understand. This is just to note that the phenomenon of deference is a distinctive response to self-ascriptions that are made from the first-person point of view.

Bar-On seems to recognize that this is the case. Consider the following passage:

> Speaking my mind is something I am in a unique position to do. Only I can express, or give voice to, *my own* present states of mind; and it is only states of my *mind* that I can express… (2004, pg. 337)

But the issue is not really whether some other person could "speak my mind", but why it is that I can only do it some of the time--only when I am related to my mind in a first-personal way. When I relate to my psychological states in more indirect or third-personal ways, they are not "expressible" and I cannot "speak from" them. I may nevertheless self-ascribe psychological states that I am detached from and I may do so correctly. That is, I may, on the basis of my therapist's suggestions, correctly say "I believe that my brother betrayed me". But, as we have seen, utterances spoken in these cases do not warrant any deference. Bar-On herself recognizes this and says that "our special first-person privilege extends only to some aspects of our present mental lives," (2004, pg. 338) But, taking the first-personal character of deference seriously amounts to explaining why that is the case, not merely noting that it is.

---

[67] Bar-On attempts to give a picture of what such an explanation would look like in Chapter 8 of her book. It is not clear to me, however, whether or not she endorses the account or is just providing a narrative of what such an account would look like.
[68] cf. Williams (2002), Ch. 5.

David Finkelstein's expressivism does attempt to take the first-personal character of deference seriously. He argues that psychological self-ascriptions are "directly expressible" only when they involve conscious states:

> Like all mental states, the unconscious ones may be expressed in our behavior. But what's distinctive about unconscious mental states is that we are unable to express them by self-ascribing them." (2003, pg. 119)

The property of consciousness is supposed to explain why certain psychological states are expressible and others are not. On Finkelstein's view, we fundamentally cannot express an unconscious mental state, even when we self-ascribe it. But why is that? What is it about consciousness that allows a psychological state to be expressed? One could say that consciousness is a kind of awareness or that consciousness states are directly expressible because they are known.[69] But this would make our practice of deferring to psychological self-ascriptions too much like the purely epistemic kind of deference. Finkelstein could always say that it is a brute fact about the world that only conscious states can be directly expressed. But this line of explanation is *ad hoc* and also probably incorrect. By emphasizing consciousness, Finkelstein is, I think, making a point similar to one that I have been making. He is claiming that it is only when a person relates to her psychological states in a distinctive way that her self-ascriptions are entitled to deference. On this point, he and I are in full agreement. I, however, do not think that consciousness is the best way to characterize what is distinctive about relating to one's own psychological states in an engaged first-personal way. Instead, I think we should focus more closely on the distinctive role that agency plays in having psychological states in this first-personal way. The first-person point of view is fundamentally the agent's point of view.

III

In the previous chapter, I suggested that we interpret first-person authority in an agential sense, as the type of authority a person has by being the only one able to do something with respect to her own psychological states. I argued that a person has this kind of authority when she can determine what her psychological states are directly on the basis of reasons. As an example, consider my belief that El Cerrito is

[69] Finkelstein himself would not say this. One of the virtues of his account is that he does not simply think of consciousness in epistemic terms. Thus, on his view, a conscious state is not directly expressible in virtue of being a state its owner is aware of.

a nice place to live.[70] Regardless of who knows better whether I believe this, I am the only person who can change this belief on the basis of good reasons. Since it is the nature of a belief to represent the truth, good reasons for believing will necessarily be truth-conductive reasons. But, even if you have a very good grasp of what is in fact true about living in El Cerrito, that fact will not change my belief. Anything you could do would only change my belief indirectly, through some kind of causal mediation. Only I can constitute my belief that El Cerrito is a nice place to live directly by judging that I ought to believe it on the basis of my sense of good reasons to believe things. This is a capacity for a distinctive type of cognitive agency with respect to my beliefs that other people necessarily lack. It is why I have a special kind of authority for my beliefs. Normally, a person has this capacity and the agential authority that comes with it simply by believing things, by having beliefs in the ordinary first-personal way. When I am not alienated or disengaged from my beliefs in some way, there is nothing further I need to do to acquire this agential authority; I have it in virtue of being a rational believer.[71]

When I believe things in this authoritative way, I sometimes tell others about them. But, in order to sincerely and competently engage in a speech act that self-ascribes a belief, a person must have some grasp of the concepts that she is self-ascribing. Because a belief is a state whose existence ought to be exclusively determined by the truth it represents, a person cannot have the concept of belief unless she understands this basic relation between belief and truth. So, I, as someone who talks about my beliefs, must understand that they are states that should be responsive to the truths that they represent. I must also understand that they as psychological states that should be changed in the face of relevant reasons indicating that they are not true. Because this is a fundamental aspect of believing, it is not something that a competent speaker can be ignorant of while sincerely self-ascribing a belief.[72]

---

[70] See also Chapter 2.

[71] Beliefs are, of course, only one type of psychological state. If first-person authority holds, as it is appears to, for every type of psychological state we will have to have analogous to determine other psychological states. The account for other types will be roughly the same; i.e., a person will have agential authority in virtue of a capacity to determine what her psychological states are on the basis of reasons. There will be differences for states, like desire, in that relevant reasons will not be truth-conductive or evidential reasons, but the general account will be the same. I will continue to focus on belief because the case provides us with a paradigm for how to think about first-person authority and agency.

[72] The awareness is not usually, nor need it be, explicitly represented. In saying that a person must be aware of the commitments involved with believing a certain thing, I do not mean the absurd idea that he must have explicit propositional knowledge of this.

Because this conceptual understanding is a necessary condition on sincere self-ascription, when I say "I believe that El Cerrito is a nice place to live," I present myself as someone who takes it to be true that El Cerrito is a nice place to live and as someone committed to the truth of what I believe. I present myself as a responsible believer, as someone who would change what I believe in the face of compelling reasons indicating that it is not true. I do not present these things explicitly by talking about either my capacity to immediately affect what I believe on the basis of my sense of good reasons or about my agential authority. Rather, these things are expressed by my self-ascribing a belief in the ordinary first-personal way. Having the capacity to directly determine what I believe for reasons, allows me to speak *as* an authority whenever I talk about by own beliefs. My agential authority is thereby *expressed* when I say something about my own psychological states.

Whenever a person is in a position of authority with respect to a specific domain they speak as an authority when they talk about facts in that domain. When one speaks as an authority she expresses her authority. Consider again an umpire at a baseball game. He has a kind of agential authority because he is the only person who can determine whether or not a pitch is a strike. He need not ever speak about his authority for this to be true. Nevertheless, when he says "that pitch is a strike", he expresses his distinctive authority to determine that the pitch is a strike. Similarly, if I have agential authority on my beliefs, I speak as authority on them and express that authority in my belief self-ascriptions.

It would be reasonable for one to wonder why speaking about psychological states should matter to the practice of deference. Why does deferring seem to require that people say things about their own psychological states? This is a significant question because it seems clearly possible for a person to have psychological states without having to speak about them, without uttering self-ascriptions. She could, for instance, say things like "it is raining". This would sufficiently express her beliefs about the rain but it would not result in a proposition that listeners would be justified in deferring to.[73] As I have argued, a person has agential authority in virtue of believing things in the ordinary first-personal way. So, she need not reflect upon her beliefs or talk about them in order to have this special kind of authority. It therefore seems reasonable to think that a person expresses agential authority by simply saying what she believes, for example, by saying "El Cerrito is a nice place to live". But, if we can express our authority without talking about our psychological states, then why is it only self-ascriptions are entitled to deference? It cannot be because an expression of one's authority is sufficient to warrant the deference of

[73] This is not to say a listener would not be justified in believing the proposition. In many cases she will. But whether someone should believe an assertion of "it is raining" depends on epistemic considerations. The specific kind of deference we are trying to understand is not warranted in this case.

others. The expressive character of a psychological self-ascription can only be part of the explanation for the phenomenon of deference.

IV

Even though a psychological self-ascription is not the only way a person expresses her agential authority, it is the only way she makes *explicit* what she takes herself to believe. [74] By talking about my own beliefs, I speak about both myself and my psychological condition, which I do not do when I talk about the weather or about El Cerrito. Thus, it is only by talking about my psychological states that I semantically represent them to others in the content of what I say. In conjunction with a speaker's expression of authority, this semantic feature of her psychological self-ascriptions helps to explain why others are entitled to defer to them.[75]

I do not have any authority over the weather; so when I say "it is raining", there is nothing that I can do, no agency I can exercise, that will make a difference to whether or not it is raining. Even if my agential authority is expressed in what I say, I have no authority over what I'm talking about because facts about the weather are outside of my control. But, things are different when I say "I believe it is raining". In this case, I semantically represent a fact over which I do have authority. Therefore, a listener is entitled to presume that what I say is true *because* what I say semantically represents the very fact that I alone have authority to determine. This response to an authoritative self-ascription is analogous to a listener's response to an umpire calling out "this pitch is a strike". The umpire's claim is presumed to be true not because it is very good evidence for the pitch being a strike. Rather, it is because listeners take the umpire to be the authority for whether the pitch is in fact a strike (this is true even when the umpire is wrong--a player or manager may argue with the umpire but he is the only one who can overturn his decision in the face of counterevidence). If one is entitled to take the umpire to have this kind of authority, one is also entitled to presume that what he says about pitches is true.

---

[74] The importance of the speaker's take on his own beliefs is connected to her own sense of her responsibilities and commitments and this is important even when her take on her beliefs is false, when it is not an authoritative psychological self-ascription. This is something Moran (2005) has suggested matters a great deal to testimony. But it is not a topic I will address here.

[75] The sophisticated expressivist was therefore correct to notice that the semantic representation of a psychological state is necessary for explaining deference. On this point, however, it seems that most philosophers theorizing about the phenomenon of deference would agree. The semantic content of a self-ascription is, after all, what we are presuming to be true when we defer.

Similarly, if one is entitled to presume that I am an authority for my own psychological states, one is entitled to presume that my sincere self-ascription of a psychological state is true. Therefore, two features of an ordinary psychological self-ascription come together to account for the phenomenon of deference. First, because a person has distinctive capacities for cognitive agency in virtue of relating to her psychological states in an engaged first-personal way, she expresses her distinctive agential authority whenever she self-ascribes a psychological state. Any listener who takes her to be an ordinary psychological subject recognizes this expression of authority.[76] Secondly, self-ascribing a psychological state also semantically represents a fact over which a speaker uniquely has agential authority. For this reason, any listener who hears the speaker expressing her authority is rationally entitled to presume immediately that what she says about her psychological states is true. Conversely, if a listener did not immediately presume that a psychological self-ascription was true, she could not see the speaker as having first-person agential authority.

This account vindicates a thread in the earlier discussion of expressivism. The expressivist was right to think that the expressive properties of an assertion, not their epistemic properties, explain the practice of deference. But the crucial thing being expressed is not an underlying psychological state, but the speaker's authority, as a cognitive agent, to determine what her psychological states are. Expressing agential authority for one's own psychological life is a fundamental part of being a person with psychological states. Part of the reason that we treat a person's psychological self-ascriptions as true is because we rightfully expect the speaker to possess the capacity to determine what the underlying psychological state is. Also vindicated is the traditional idea that deferring is intimately connected to first-person authority. As many philosophers have thought, it is because a person is a special kind of authority with respect to her own psychological states that we justifiably defer to what she says about them. But, if what I have said is correct, it is not because the person has any kind of epistemic authority.

Having a special kind of authority to directly constitute one's psychological states on the basis of reasons is central to having psychological states in the way that we do; it is fundamental to being a person with a psychological life. One cannot doubt that a human being lacks this authority without at the same time doubting that she has a psychological life to speak about. To the degree that we view someone as speaking about her own psychological life we *ipso facto* credit her with normally having this kind of authority. Once we do so we are entitled to defer to what she says about her psychological states. As I said early on, this is a presumption and the presumption may be overridden at times for reasons both familiar and unusual, as,

[76] This is not to say that they recognize it *as* an expression of authority. People normally have, express, and respond to agential authority without conceiving of it as such.

for instance, when I self-ascribe psychological states on the basis of therapeutic interpretations or observations of my behavior. But the presumption cannot be completely dissolved. We cannot completely stop deferring to what a person says about her own psychological states without thereby ceasing to view her as someone who is a cognitive agent capable of taking up the first-person point of view.

V

If the practice of deferring to what a person says is essential to treating them as a cognitive agent, then any such subject should be entitled to deference. But what would happen if a third party were in an especially good position to know about my psychological states? What if their knowledge were as direct and reliable as my own or even better? This would appear to make them into a kind of epistemic expert on my psychological life and that would mean they should be entitled to a purely epistemic kind of deference. Could this person's epistemic expertise be so thorough that others should defer to what she says about my psychological states over what I do? These questions were raised at the beginning of this chapter and, since we now have a proposal for understanding what justifies the practice of deference, we should return to them.

The thought behind these questions is that the practice of deferring to a person's psychological self-ascriptions might be superseded by a purely epistemic kind of deference. If some other person were to have especially good epistemic access to my own psychological states, it seems reasonable for others to defer to what they say about my beliefs, desires and feelings. We might even try to imagine what would happen in a possible world where some other person had better epistemic access to my psychological states. Would the practice of deferring to my self-ascriptions simply not exist in such a world? Even if I am right in thinking that the epistemic properties of assertions are not why we defer, perhaps certain epistemic parameters are a necessary background constraint on justified deference. If so, then being entitled to deference would not, as I said, be fundamental to being a person with psychological states; it would instead be a contingent entitlement, one that listeners could get around in certain epistemic situations.

This possibility was discussed by A.J. Ayer in his paper "Privacy". Ayer asks us to consider the possibility of what he calls co-consciousness:

> But suppose that someone did claim to be 'telepathically' aware of some other person's mental states and that his reports of them were found to be consistently true: and suppose that when he was asked how he did it he said that it was like asking him how he knew what his own thoughts and feelings were; so far as he was concerned the only difference was that the other

> person's thoughts and feelings came to him under a different label. (1963, p.p. 66-67)

Ayer doesn't give us too many details for how to imagine this scenario. Some philosophers might reject the example because they believe that any kind of consciousness of a psychological state entails ownership of it. But there is no obvious reason to think that this entailment holds.[77] There does not seem to be anything contradictory to the idea that another person could have some kind of epistemically direct access to my psychological states. So, for the sake of argument, let's suppose co-consciousness is possible. It would seem to have properties similar to ordinary consciousness. Like ordinary consciousness, co-consciousness would be epistemically direct or non-inferential. Also, since we no longer think that consciousness is infallible, we should not take co-consciousness to be infallible. Just as I may make mistakes about my own mind, another person who is co-conscious of psychological states may be mistaken about their character. This last point is important because fallibility means it is possible that someone who is co-conscious of my psychological states may in fact be in a better position to know about them than I am.

Suppose that Jane is co-conscious of my psychological states. Usually, when I think, feel or desires something, Jane knows that I do in an epistemically direct way. Suppose further that over time Jane becomes especially good at discerning precisely what my psychological states are. She may develop this expertise to the point where she is in a better position to know about them than I am. Perhaps I suffer from some sort of neurological defect or am so overwhelmingly busy with work that I cannot focus carefully enough on some of my more complicated psychological states to know about them. Conversely, Jane might spend most of her time thinking about the exact character of each of my psychological states and, because her brain is in perfect health, her faculty of co-consciousness might function extraordinarily well. If something like this were the case, Jane would be in the best position to know what my psychological states are. This would be clearly indicated by the fact that her pronouncements were less subject to error and far more reliable than anyone else's, including my own. From a strictly epistemic point of view, it would seem that Jane is

---

[77]Another reason to think the entailment does not hold is schizophrenia. One characteristic of the schizophrenic delusion of thought insertion is the presence of "alien" thoughts to which one has epistemically direct access. Campbell (1999), I think rightly, claims "the content of the schizophrenic's illusion is that he has first-person knowledge of token thoughts which were formed by someone else. And there is no immediate contradiction in that." (pg. 620) There would be a contradiction if the entailment from consciousness to ownership held but then, I think, we could not achieve a proper understanding of thought-insertion. I will discuss this phenomenon in a bit more detail in Chapter 5.

the best person to ask if you wanted to find out what I was thinking or feeling. I admit that this example is far-fetched but it brings out the relevance of epistemic issues to our ordinary practice of deferring to psychological self-ascriptions.

Would a listener be justified in continuing to defer to what I say about my psychological states over what Jane says? Ayer thinks that she would. Of this example, he writes that "even if we allow it to be possible for others to become aware of his thoughts and feelings in the way that he does, their knowledge of them will be subordinate to his." (1963, pg. 68) But why is this true? Why would Jane's assertions about my psychological states be subordinate to my own when, *ex hypothesi,* her claims are more likely to be correct? If she is in a better position to know about my psychological states, it would seem to make more sense that my knowledge would be subordinate to hers. Ayer cannot be claiming that Jane's assertions about my psychological states are epistemically subordinate to my self-ascriptions. That is simply not the case. If we were only interested in knowledge of facts, what Jane said about my psychological states would be entitled to deference more than what I did.

Nevertheless, Ayer believes that Jane's assertions are subordinate because, when it comes to my psychological states, I am "the final authority concerning their existence and character." (1963, pg. 68) It is important to notice Ayer's point is that even if Jane has a kind of epistemic expertise, she is not the authority on the "existence and character" of my psychological states. I think what Ayer recognizes is that even when another person's attributions of psychological states to me are not *epistemically* subordinate to my own, the entire practice of deferring to epistemic authorities like Jane is always subordinate to the distinctive kind of deference we ordinarily show psychological self-ascriptions. We defer to a person's psychological self-ascriptions because we rightly take the speaker to have a unique capacity to determine the very facts that she is talking about. But these are also the facts that someone like Jane knows about. So her epistemic expertise, no matter how thorough, cannot supersede the authority that determines the "existence and character" of facts in the domain over which she is an expert.

When Ayer claims that a person is the "final authority" on the "existence and character" of her own psychological states, we should understand this "authority" in an agential sense. In that sense, he is right to think that a person's authority cannot be superseded. By contrast, epistemic authority of the sort we are supposing Jane to possess is contingent. Were an accident to suddenly change things, she would no longer be in the best to know about my psychological states. But, even though I fail sometimes, it does not seem that I could entirely lack the capacity to determine what I believe or desire directly on the basis of reasons. If I did there would be a compelling reason to think that I simply did not have beliefs or desires.[78] So, even

[78] To take just one example, consider how, in order to change someone's mind, we normally attempt to persuade a person with reasons. If we did not presume that her

though a person may not possess agential authority for every one of her psychological states--she may suffer from an occasional neurosis or other type of cognitive pathology--it does seem she must have this special kind of authority for most of them. For this reason, I do not believe any measure of contingent epistemic authority could supersede someone's agential authority and the entitlement to deference it brings. Cognitive agency is an essential part of our nature as persons with psychological states, which means that its authority is also essential to having our type of psychological life. If this is right, then deferring to a person's psychological self-ascriptions is not merely a way of being polite or a default courtesy we tend to practice; it is an essential part of treating the individual who is speaking to us as a person.

---

judgments about what she ought to believe or desire would immediately affect what she did believe or desire, such rational persuasion would have no point. But, if this were the case, it would also not at all be clear that the person had beliefs or desires. This point is discussed in much more detail in the following chapter.

CHAPTER 4
SELF-BLINDNESS AND SELF-AWARENESS

In order for me become aware of another person's psychological states, I must observe the person in some way. I must see what they are doing or listen to what they are saying before I can know what is on their mind. Even when I learn about a person's psychological states indirectly from someone's testimony, it is because another person observed her behavior. Because observation is, either directly or indirectly, necessary for becoming aware of another person's psychological life, I can be, and often am, mistaken about what others are thinking or feeling. Sometimes my perceptual system can generate errors even when I am rational in how I form my beliefs. The possibility of this particular kind of error demonstrates that another person's psychological state is ontologically independent from my own state of knowledge of it; the two are, in Hume's words, distinct existences.

In my own case, however, I do not typically have to make observations in order to know what I am thinking or feeling. Instead, I seem to have a special kind of epistemic access to my own psychological states. Because this access is, in principle, unavailable to others, it is often called first-person access. But, my having first-person access makes it far less clear whether my own psychological states are ontologically distinct from my knowledge of them. The phenomenon of first-person access can easily suggest that something stronger than a causal relation connects an individual's psychological states with her own knowledge of them. In fact, to many philosophers it has seemed that, within a single individual's mind, the two are not distinct existences but are ontologically interdependent.[79]

This last idea is explicitly endorsed, for example, by Akeel Bilgrami: "our mental states lack a certain kind of independence from our knowledge of them…possessed by the things *of* which we have perceptual knowledge, *from* that perceptual knowledge." (2006, pg. 29)[80] Bilgrami suggestively claims that a person's knowledge of her own psychological states is "not come by via anything at all…it comes with the states of minds themselves, for 'free', as it were." (2006, pg. 38) The basic idea behind this metaphor of "knowing for free" is that a psychological state is not ontologically distinct from its owner's knowledge of it. For this reason, no substantial account of the owner's knowledge is needed. In fact, were we to attempt to offer a detailed theory of first-person access, we would significantly mischaracterize the phenomenon. For a view like Bilgrami's, the ontological dependence of a person's psychological states on her states of knowledge is sufficient to explain her distinctive first-person access to them.

---

[79] There are various ways to make out the interdependence. In addition to Sydney Shoemaker's view, which I focus on in this chapter, Bilgrami (2006) and Heal (2001) both endorse a form of ontological interdependence.

[80] cf. Bilgrami (2010).

Opposing this entire line of thinking is a natural and very simple thought; namely a psychological state is one thing and having knowledge of it is something else. We saw that this is to true for our knowledge of psychological states belonging to other people. Why is it not equally true for our own? If the two cases are symmetrical, then a person's psychological states and her own knowledge of them would also be ontologically independent; they would be distinct existences. I propose to capture this idea with the following thesis:

Distinct Existence Thesis: For any subject S and psychological state *M*, ~ (S has *M* only if S know that S has *M*).[81]

Intuitive considerations speak in favor of the *Distinct Existence Thesis*. First many of a person's psychological states represent mind-independent material facts that have nothing to do with her own psychological life. But, a person's beliefs about her own psychological states represent only facts about her mind. Assuming, plausibly, that a psychological representation is sensitive in some way to whatever fact it represents, then since most material facts are clearly ontologically independent from psychological facts, it might seem, *prima facie*, that representations of the former could exist without representations of the latter.[82] Secondly, we are both ignorant and mistaken about our psychological states. It seems that we can believe or desire something without knowing that we do and that we can mistakenly self-ascribe feelings, desires and beliefs that we do not actually have. How does this happen? Why do we make errors about our own psychological lives? Ignorance and error in this domain strongly suggests that the *Distinct Existence Thesis* is true. Even if errors are thought to be quite exceptional, one can reasonably expect some account of why or how they occur when they do. The most plausible theory of error will appeal to

---

[81] A kind of ontological dependence is also a consequence of an infallibility thesis about self-knowledge. I shall not discuss this sort of thesis in this chapter because few people believe we are infallible about our own psychological states. However, if, as I shall argue, all psychological states are distinct existences, they are ontologically independent from beliefs about them as well. In that case, we would need to broaden the *Distinct Existence Thesis* to a biconditional and phrase it terms of belief rather than knowledge: For any subject S and psychological state M, ~ (S has M iff S believes that S has M). This is because all knowledge depends (trivially) on what is known.

[82] This point is related to familiar discussions of Moore's Paradox. The conjunction (P and I believe ~P) is not a logical contradiction because each conjunct is an independent fact. Does Moore's Paradox occur at a higher level? Consider the following conjunction: (I believe that it is raining but I believe that I do not believe that it is raining). One who denies the *Distinct Existence Thesis* could claim that this sort of conjunction did entail a contradiction on the grounds that the two conjuncts were not independent.

the ontological independence of psychological states from our knowledge of them. It will maintain that mistakes occur because a contingent relation connecting a person's psychological states with her knowledge of them fails to hold.

Despite these sorts of considerations, many philosophers resist the *Distinct Existence Thesis* and endorse views similar to Bilgrami's. This resistance is often supported by an appeal to a well-known argument presented by Sydney Shoemaker in a number of papers. Shoemaker argues that if psychological states were ontologically independent from our knowledge of them an absurd condition he calls "self-blindness" would be possible. A person suffers from self-blindness just in case she can know about her own psychological states in only a third-personal way.

In this chapter, I shall defend the *Distinct Existence Thesis* from Shoemaker's argument. I will argue that even if self-blindness is impossible the best explanation of this impossibility does not deny that psychological states and our knowledge of them are distinct existences.

Much of Shoemaker's own resistance to the *Distinct Existence Thesis* is motivated by his recognition that first-person access is epistemically unique. If the *Distinct Existence Thesis* is true, it might seem that our way of knowing about our psychological states would be grounded in causal relations. In that case, however, it would no longer be distinctive; it would be too much like perceptual knowledge. In the final section of the chapter, I shall offer an account of the epistemic distinctiveness of first-person access without appealing to the ontological status of psychological states. On the view I propose, first-person access is fundamentally connected to our natures as cognitive agents rather than to the ontological status of psychological states. If this is right, we can accept the *Distinct Existence Thesis* without undermining the epistemic distinctiveness of first-person access.

I

According to Shoemaker, "a self-blind creature would be one which has the conception of the various mental states, and can entertain the thought that it has this or that belief, desire, intention, etc., but which is unable to become aware of the truth of such a thought except in a third-person way." (1988, pg. 115) Self-blindness is supposed to be analogous to color-blindness. Just as a person who is color-blind can learn information about color in a non-standard way, through reading a book about colors for instance, a person who is self-blind should be able to learn about her own beliefs, desires, and intentions in a non-standard way, which is to say a completely third-personal way.[83] Instead of having first-person access to her own

[83] Other important accounts of this kind of possibility can be found in Bilgrami (2006) and Peacocke (2008), although neither draws the same conclusions as Shoemaker.

psychological life, the self-blind person will have to make observations of her behavior in order to know what she believes, wants or feels. Moreover, just as the color-blind person does not suffer from a cognitive or conceptual deficiency, self-blindness is "supposed to be perceptual or quasi-perceptual, rather than cognitive or conceptual." (1988, pg. 118) Shoemaker argues that if the *Distinct Existence Thesis* is true, self-blindness should be possible. It should be possible for someone to have the same kinds of psychological states any of us do without having any of the normal, first-person access to them that we do.[84]

Shoemaker's argument is a simple. Although he tailors the argument to many different types of psychological states, each variation rests on a thought experiment asking us to imagine a self-blind person with respect to a specified type of psychological state.[85] I will focus on the case of beliefs but the argument is roughly analogous for other psychological states. When he discusses beliefs, Shoemaker asks us to try to imagine a self-blind man he names George. If George were conceivable, if we could successfully picture someone learning about his beliefs in only a third-personal way, the *Distinct Existence Thesis* might be plausible. But, Shoemaker argues that we cannot really imagine such a person.

To bring out the incoherence of self-blindness, Shoemaker's focuses on Moore's Paradox. If George were self-blind, we should be able to imagine that "the total evidence available to a man at a given time should support the proposition that it is raining, while the total 'third-person' evidence available to him should support the proposition that he does not believe that it is raining." (1988, pg. 118) In this case, it seems reasonable for George to assert a Moore-paradoxical proposition like "It is raining, but I do not believe that it is raining," because it is what his evidence supports.[86] But, Shoemaker quickly points out that a rational person recognizes that

---

[84] cf. Shoemaker (1994)

[85] The different versions can all be found in Shoemaker's Royce Lectures (1994).

[86] There is another puzzle here about Shoemaker's example. He claims that the total third-person evidence might justify the assertion of a Moore-paradoxical sentence. Offhand, it may seem that this evidential situation could arise because there are two distinct facts involved; one about George's psychology and one about the weather. However, I think that when one really tries to imagine the situation it becomes nearly impossible to picture the totality of George's evidence unambiguously supporting both conjuncts of a Moore-paradoxical proposition. Suppose we include George's utterance of "it is raining" as one component of the "third-person" evidence. We will no longer have an evidential situation that clearly supports George asserting "it is raining but I don't believe that it is raining." Shoemaker oddly seems to grant this at one point, but says that "since the objection under consideration" depends on an inconceivable situation (unambiguous evidence for a Moore-paradoxical proposition) and his argument does not, the objection is defeated. But, the problem is that it looks like Moore's paradoxical assertions may be avoided because of objective, third-party,

Moore-paradoxical assertions are inappropriate. Since we are supposing that George is rational and not conceptually or cognitively deficient, he will avoid asserting a Moore-paradoxical proposition, thereby behaving in the same way any of us would.

Moreover, because George has the concept of belief, Shoemaker thinks the following two things will be true:

(A) He will recognize that when asked "'Do you believe that P?' he ought to answer 'yes' just in case he would answer 'yes' to the question "Is it true that P?"

(B) He will recognize the meaning of 'believe' and "preface his assertions with 'I believe' in just the circumstances in which this is pragmatically appropriate.

Plausibly, (A) follows from George's being a rational believer with the concept of belief.[87] When asked whether or not you believe that P, you, like George, typically consider the truth of P.[88] The concept of belief is that of a state responsive to the truth it represents. Understanding this connection between belief and truth is a large part of having the concept of belief. Since George has the concept, he will plausibly act like the rest of us when faced with questions about his beliefs.

Having the concept of belief also means George can appreciate the relevance of pragmatic considerations for self-ascriptions of belief. For example, when I lose my keys, I have to look for them. Where are they? I don't really know and there is not enough evidence indicating their location. But, I *believe* they are in my office. In

---

evidence. If that is true, it is unclear how the phenomenon helps us understand first-person access.

[87] I actually have some doubts about whether (A) is true for George. Shoemaker thinks that the truth of (A) is central to our concept of belief. But, it seems to me that part of why ordinary believers can answer the question "Is P true?" when asked about their beliefs is because they can presuppose that they have first-person access to their cognitive actions. If I did not know what I *judged* when answering "Is P true?" it is hard to see how I would be rational in considering "Is P true?" as a way of answering a question about my beliefs. Thus, even though (A) is true for us, it may not be *central* to our concept of belief. However, I will grant that (A) is true for George for the sake of the argument in this chapter.

[88] This is a familiar point from the Evans we considered in Chapter 2. Once again, that passage states: "In making a self-ascription of belief, one's eye's 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*." (1982, pg. 225)

this sort of case, my saying "I believe the keys are in my office" is a way of making a hesitant assertion about the location of my keys. Since he has conceptual mastery, George would behave the same way I do when I lose my keys. Like me, he could say "I believe that the keys are in my office" in appropriate contexts.

Because of these truths, Shoemaker concludes that George will behave just like any normal person and self-ascribe beliefs in the same conditions we would. There would "be nothing in his behavior, verbal or otherwise, that would give away the fact that he lacks self-acquaintance." This, however, just means that self-blindness is not really conceivable.[89] When we try to imagine George, we can only imagine a person who behaves like us, which means that we *cannot* imagine a distinctively self-blind person.

It is important to keep in mind that George is supposed to be *completely* unable to become aware of his psychological states except in a third-personal way. Even if we can conceive of him lacking first-person access to some psychological states, this is not self-blindness. After all, we lack first-person access to some psychological states some of the time, but we are not self-blind. To truly conceive a self-blind person, we need to try to imagine a person as sophisticated as we are psychologically and conceptually without any first-person access at all. Is this possible?

I agree with Shoemaker that it is not. It is certainly difficult for me to imagine a fully rational person with a psychological life as rich as an ordinary adult human without that person having first-person access to her own psychological states. Such a person would *always* have to observe her own behavior to learn about her psychological states but then it can seem that she might have no psychological life at all. Without a prior grip on her own thoughts, why would she take her own behavior to be intentional behavior? Why would she not instead take it to be a series of involuntary movements? Since this person has no first-person access to her beliefs, or desires, it does not seem she has any reason to take herself to be a cognitive agent with psychological states rather than an automaton, presuming that she can *take* herself to be anything.

Although I personally find these considerations compelling, I don't believe they are conclusive. In fact, I think one might try to make a case for the conceivability of self-blindness by imaging someone losing first-person access later on in life, after acquiring a great deal of rational maturity and conceptual mastery. Just as it is difficult to imagine someone blind from birth mastering color concepts, it is hard to picture someone who has always been self-blind with the resources necessary to self-ascribe psychological states. In both cases the difficulty might be

---

[89] cf. (1994), pg 233: "If self-blindness is possible with respect to a given mental phenomenon, one would think that there should be something that would show, or at any rate provide good evidence, that someone was afflicted with such self-blindness."

overcome by imagining the person losing the specific mode of access as an adult. If we try to imagine George slowly losing his capacity for first-person access, might we not have more success? Although I think we might and it might be instructive to try, I still find self-blindness incoherent. Thus, for the remainder of this chapter, I shall assume that it is impossible in order to see what this means for the *Distinct Existence Thesis*.

II

Recall that the *Distinct Existence Thesis* embodies the rather simple idea that psychological states are ontologically independent from our states of knowledge of them, that having a psychological state is one thing and knowing about it another thing. If this is so, it will be possible for any particular psychological state to exist without my being aware of it in the ordinary first-person way. Shoemaker's argument from self-blindness is taken by many philosophers to show that this is impossible. He himself characterizes the target of his argument as a view that holds "the existence of these states and events is independent of their being known in this way, and even of there existing the mechanisms that make such knowledge possible." In describing his target this way, however, Shoemaker conflates two distinct ideas. First, there is a possible dependence of the existence of a psychological state on the existence of another, namely the state of knowing the first. But, secondly, there is a possible dependence of the existence of a psychological state on what Shoemaker calls "mechanisms that make such knowledge possible", which we can think of more simply as the capacity for first-person access to one's own psychological states. It is helpful to distinguish these two.

State Dependence: For any subject S and psychological state *M,* necessarily, S has *M* only if S knows that S has *M*.

Capacity Dependence: For any subject S and psychological state *M*, necessarily, S has *M* only if S has a distinct capacity to know that S has *M*.

Because Shoemaker intends for his view to be compatible with recent evidence in social psychology indicating that we sometimes lack awareness of our own thoughts and feelings, *State Dependence* should be understood as a generic claim.[90] It is also a

[90] It would therefore strengthen *both* to read the claims universally. But, I assume that the stronger version of either would be unacceptable to most contemporary philosophers. However, the stronger form of *State Dependence* does seem to have been Descartes' view.

stronger claim because *Capacity Dependence* follows from it trivially. Nonetheless these two dependence relations are logically distinct.

The *Distinct Existence Thesis* is clearly inconsistent with *State Dependence*. If psychological states are ontologically independent from our states of knowledge of them, a person could have psychological states without knowing it. This is even how Shoemaker at one point describes self-blindness: "that for each kind of mental fact to which we have introspective access, it is at least logically possible that there should be creatures in which such facts obtain, and who have the ability to conceive of them, but who are self-blind with respect to them."(1994, pg. 273) Although this is a consequence of the *Distinct Existence Thesis*, it is a much *weaker* characterization of self-blindness than the standard one we find in Shoemaker's discussion of George. Even if it is possible for a creature to be self-blind with respect to "each kind of mental fact", it may nevertheless be *impossible* for the same creature to be self-blind with respect to all of them.[91] Therefore, although a defender of the *Distinct Existence Thesis* must deny *State Dependence*, this alone does not commit him to the possibility of complete self-blindness.

When describing George, however, Shoemaker claims he is unable to become aware *any* psychological states except in a third personal way. If George were self-blind to every psychological state, he would lack a *capacity* to know about his psychological life in the ordinary first-person way. If Shoemaker's argument is correct, this kind of self-blindness is impossible. A person could not have a psychological state in the same way that we do without having a capacity for first-person access to it. This would mean that *Capacity Dependence* is true. But what makes it true? Why must someone with our kind of psychological states have a capacity for first-person access? The simple answer to these questions is that *State Dependence* is true. If my psychological states were to depend on my knowledge of them, it would follow that I have a distinct capacity for knowing about them, one grounded in the ontological dependency. However, this would require that we give up the *Distinct Existence Thesis*, which is something we may not have to do if we can explain *Capacity Dependence* in another way.

---

[91] Consider a slightly imperfect analogy. A person being a student is independent from her receiving an 'A' in a Philosophy 101. Thus, for any student in this class, it is possible that she does not receive an 'A'. The two states are ontologically distinct. But it may nevertheless be *impossible* for no student to receive an 'A' in this course. Suppose, plausibly, that if it began to look as if all the students in the course were going to receive less than an 'A', the grading scale would necessarily be recalibrated (it could be a part of what it is to be a introductory university course). Thus, it is impossible for *every* student to not receive an 'A', although it is possible for *each* student.

III

The notion of first-person access is intended to capture the fact that each of us has a distinctive way of knowing about our own psychological states. We all seem to have epistemic access to them in a special, first-personal way, unavailable to others. Remarking on this, Donald Davidson pointed out that "it is clear that people normally know without observation or evidence what they believe, want and intend."[92] The basic phenomenon is just what Davidson describes--a way of knowing that is not observational and evidentially groundless. Whenever a person knowingly self-ascribes a psychological state, they seem to do so in virtue of having that state in the ordinary first-personal way. But, crucially, having the state in this first-personal way is not the person's *epistemic basis* for knowing that they have it. Although it is tempting to characterize first-person access more substantially, as, for instance, something incorrigible or infallible, I want to focus on this basic phenomenon, which I think is relatively uncontroversial.

Even if we assume that Shoemaker is right and that having psychological states requires a capacity for first-person access to them, this is not obviously due to the nature of psychological states. Instead it may have something to do with peculiar features of persons *qua* psychological subjects. Shoemaker himself sometimes writes as if he has this sort of thing in mind. For instance, he seems to endorse the idea "that it is of the essence of mind that each mind has a special access to its own contents, or more soberly expressed, that each person has a special access to his own mental states."(1988, pg. 115) But if "the essence of mind" includes this "special access", couldn't "the essence of mind" account for *Capacity Dependence*? If so, it does not seem we would need to appeal to *State Dependence* to explain this "special access". We have already seen how difficult it is to imagine a person without first-person access. Perhaps this is because having first-person access is part of what it is to be a person.

A central aspect of a being a person is being a rational cognitive agent whose beliefs, desires, intentions and other psychological states are responsive to appropriate features of the world. My beliefs, for example, are psychological states that are normally sensitive to evidence and to reasons bearing on the truths they represent. Thus, if I believe that P and am confronted with what I take to be evidence that P is false, I will, if I am rational, immediately stop believing P. In this way, my own assessment of the world, my take on evidence for or against P, makes an immediate and substantial difference to the existence and character of my psychological state. As a subject with psychological states that are responsive to

---

[92] Davidson (1982), pg. 100. Philosophers often try to capture what is unique about first-person access in terms of it being "direct". But, on some understandings, perceptual access could also be seen as direct. It is therefore useful to avoid the term.

reasons, my own sense of good reasons for or against having them directly affects them and, in many cases, changes them.

Moreover, from my point of view, my own beliefs and desires must be determined *exclusively* on the basis of reasons. If they were not fully determined by my own assessment of reasons for having them, if, after judging that I ought to believe that P, I had to do something else to get myself to believe that P, I would no longer function as a rational cognitive agent. If my judgment that I ought to believe that P did not settle what I believe, no one, including myself, could rationally persuade me to believe that P. In that case, the only way to get me to believe anything would be through brute force and, although this could succeed, the existence of my beliefs would not depend on only what I judge I ought to believe. To the extent that my judgment about what I ought to believe had any relevance, it would be as much as any of the other instrumental means required for changing my beliefs. But, for a rational believer, recognizing that one has a good reason for believing that P is sufficient to make it the case that she does believe that P. What a person concludes to be a good reason does not merely *seem* compelling to her but actually compels--it settles the question of what she believes.[93]

It seems to me that in order for a person's own sense of reasons to be able to directly determine her psychological states in this way she must have a special capacity for first-person access. Tyler Burge has recently stressed that first-person access is necessary for critical reflection on one's psychological states. He writes that "it is constitutive of critical reasoning that if the reasons or assumptions being reviewed are justifiably found wanting by the reviewer, it *rationally follows immediately* that there is prima facie reason for changing or supplementing them, where this reason applies within the point of view of the reviewed material (not just within the reviewing perspective)."[94] Someone reasoning critically in Burge's sense must focus her attention on her own beliefs, desires or intentions in order to assess them. This often culminates in judgments to the effect that a particular state is warranted or not by reasons, which often immediately change the person's original psychological states. Burge thinks that if our way of knowing our own psychological states rested

---

[93] cf Burge (1998): "For reasons necessarily not only evaluate but have force in forming, changing, confirming attitudes in accord with reasons. All reasons that thinkers have are *reasons to*, not merely rational appraisals." (pg. 250)

[94] (1996), pg. 257. Burge's argument is directed against what he calls the observational model of self-knowledge. Such a model allows the possibility of "brute errors" about one's own psychological states. Burge suggests that the possibility of "brute error" is partially responsible for the dissociation between the "reviewing" and the "reviewed" perspective. Although I agree that without first-person access there is a kind of dissociation, I am not convinced that the possibility of brute errors is quite the reason for it. The correct explanation for the dissociation, as I will argue, proceeds along different lines.

on some epistemic intermediary, "there would never be an immediate rationally necessary connection" between what we deliberatively judge we ought to believe and what we do believe. Why not?

In order for someone to engage in critical reasoning about her own psychological states, she would have to be aware of them. Suppose she does so indirectly or third-personally on the basis of evidence or observation. This is the same kind of epistemic access we have to the psychological states of other people and we can rationally criticize them or make judgments about what they ought to believe. Yet when it comes to others our judgments are more like appraisals or evaluations; they do not necessarily determine the other person's psychological states. Depending on how convincing we are, the other person may accept our critical judgments, but she may also easily ignore them. Whenever someone takes up this kind of third-personal perspective on her own psychological states, she treats them like facts whose existence is independent from her rational assessment. In so doing, she abdicates her rational authority for them.

In fact, because of this it seems to me that Burge overemphasizes the importance of critical reasoning. The necessity for a special kind of first-person access is really more basic. Simply having rational psychological states requires a capacity for first-person access independently of our ability to subject them to critical reflection. This is because, as rational subjects, we must take our own psychological states to be exclusively determined by reasons. When we relate to our psychological states in a first-personal way, when we are directly engaged with them prior to any critical reflection, we take their existence to depend only on good reasons for them.

Suppose that I believe that P. From my first-personal point of view, it can only be because some suitable reason R (or set of reasons) is a compelling indication of the truth of P. For me, only R matters for whether or not I continue to believe that P. However, if my way of knowing about my belief were not first-personal, if it rested on epistemic mediation of some kind, then I would have to acknowledge that my belief that P could exist without R. Basing knowledge on evidence that does not include R means admitting that the existence of my belief does not exclusively depend on R. This would directly undermine the rational authority of R, which explains why, in Burge's words, "there would never be an immediate rationally necessary connection" between my sense of what I ought to believe and what I do believe. He is right to think that such dissociation occurs without a capacity for first-person access; but it can happen prior to engaging in critical reasoning.

So, it seems that a person must be able to access to her own psychological states in a first-person way in order for them to be determined exclusively on the basis of reasons. This is not to say that she will always know about her psychological states in this way. Standing in a more third-personal relation to her beliefs or desires, she may occasionally learn of them on the basis of behavioral evidence. So having a particular psychological state does not guarantee that one has first-person access to it. When it comes to our own psychological lives, ignorance and error are both real

obstacles. But it does seem that if a person lacked the capacity for first-person access, if she lacked the ability to know what she believed or desired without evidence, her psychological life would remain at too far a distance from her sense of reasons and she would cease to be a rational cognitive agent.

First-person access is therefore tied up with the ordinary way in which each of us has psychological states. If that is true, it is impossible for anyone like us to have psychological states without having a special first-personal capacity to know that she does. *Capacity Dependence* is true but not, as Shoemaker thinks, because the existence of a psychological state ontologically depends on knowledge of its existence. Rather, it is true because being a rational cognitive agent partially consists in having a distinctive capacity for first-person access even if the existence of any particular psychological state does not. A person's rational cognitive agency grounds her epistemic entitlement to non-evidential knowledge of her own psychological states.

IV

One may wonder at this point whether there is really any substantial difference between my account of first-person access and ones, like Bilgrami's or Shoemaker's, that are grounded in *State Dependence*. On the view I am proposing, *State Dependence* is not true, but something very much like it is. *Capacity Dependence* is a modal claim to the effect that persons necessarily have a capacity for first-person access. One may easily think that we are hardly justified in saying we have such a capacity unless we did *in fact* know about most of our psychological states. That is, it might seem that in order to have a capacity to □, one must actually □ most of the time.[95] It seems difficult to understand the attribution of a largely unrealized capacity. So, it would seem that any subject of psychological states will actually be aware of most of them and, moreover, this will be necessarily true. Thus, from *Capacity Dependence*, one may plausibly infer:

SK*:* For any subject S and psychological state *M*, necessarily, *it is highly probable* that if S has *M*, S knows that S has *M*.

Because *Capacity Dependence* states a necessary truth, it might seem that we are guaranteed to know about most, even if not all, of our own psychological states. We need only actualize our capacity for first-person access, which should be rather easy.

[95] Even this, however, is questionable. See the excellent discussion in Squires (1969).

The inference to SK from *Capacity Dependence* is not obvious and needs to be spelled out a bit more carefully.[96] Having a capacity alone does not obviously require an especially high probability of successfully actualizations of it. The elevator in my building has a capacity to hold 2,500 pounds, but I doubt that it ever has. I certainly don't think that it has held 2,500 pounds frequently. The thought behind the inference to SK, however, is that capacities must be realized when the appropriate opportunities come along. It is highly probable that if I were to put 2,500 pounds into my elevator, it would hold the weight. Similarly, a person with the capacity to swim would swim when placed in water, even though they might spend most of their life not swimming at all. If such a person did not usually swim when in the water, it seems that they would lack the capacity to swim.

Two other qualifications are needed. First, we need some kind of *ceteris paribus* clause ruling out external inhibitors of the capacity in question.[97] My elevator would have the capacity to hold 2,500 even if it broke frequently because of some manufacturing defect in the cables.[98] Secondly, when it comes to an agent's capacities (as opposed to an elevator's), the agent in question must be trying to exercise the capacity in question. With these qualifications in mind, it does seem plausible that, *ceteris paribus*, if a person in water does not swim *and* they are trying to swim, they lack the capacity to swim. Analogously, it seems plausible that if, *ceteris paribus,* a person is relating to her psychological states in the ordinary-first personal way and is ignorant of them, she lacks the capacity for first-person access. In order to infer SK from *Capacity Dependence* one must also assume that it is highly probable that a person relates to her psychological states in the ordinary first-personal way. I think that both the assumption and SK are plausible so I will suppose they are true.[99]

The plausibility of this kind of inference raises the question of whether there is a substantial difference between *State Dependence* and SK. From either a pragmatic perspective or an epistemic perspective there is not much of a difference. If either is true, most people necessarily know about a great majority of their psychological states. The only difference seems to be whether an ontological dependence relation holds between a person's psychological states and her own knowledge of them. SK is compatible with the *Distinct Existence Thesis* whereas *State Dependence* is not. But what

---

[96] Ideally, one would like to make the degree of probability more precise. That, however, is beyond the scope of the argument in this chapter.

[97] There is a risk that the *ceteris paribus* clause will make the conditional trivial. This is a familiar point from Goodman's (1954). But, since SK is not intended to be an *analysis* of our capacity for first-person access, we can set Goodman's concerns aside.

[98] This is intended to rule out the masking of a capacity by some other condition(s). For more on masking, see Martin (1994) and Lewis (1997).

[99] I actually think that SK could be a datum. In fact, it seems to me that many philosophical accounts of self-knowledge are attempts to explain why SK is true.

difference does this make? Why insist that the *Distinct Existence Thesis* is true? Why not accept *State Dependence* instead?

Recall that earlier I suggested our intuitions favor the *Distinct Existence Thesis*. It should seem natural to hold the view that psychological states are ontologically independent from states of knowledge of their existence. The principal motivation for rejecting these intuitions in favor of a *State Dependence* view is to make sense of the phenomenon of first-person access. In the previous section, I presented an alternative way to understand that phenomenon, which should therefore diminish most of the attractiveness of *State Dependence*. One prefers, I presume, a hypothesis more consistent with our intuitions. But, these intuitions can also be made more precise in ways that bring out serious shortcomings with *State Dependence*.

To better understand these problems, it is useful to focus on a specific example of a *State Dependence* view. I am therefore going to describe Shoemaker's own view in slightly more detail. Shoemaker believes that a psychological state and a state of knowledge of it "have the same core realization and that the total realization of the first-order state is a proper part of the total realization of the first-person belief that one has it." (1994, pg. 243) We can understand the "core realization" of a psychological state as whatever brain state realizes it. Shoemaker thinks that the very same brain state also realizes the person's knowledge of the primary psychological state. In order to realize this state of self-knowledge, there must only be "a certain degree of rationality, intelligence, and conceptual capacity (here including having the concept of belief and the concept of oneself);" when there is, "automatically one has the corresponding second-order belief."(1994, pg. 243) This is why the two states have different "total realizations". One can lack a second-order belief even when the relevant brain state is present but only if one also lacks "a certain degree of rationality, intelligence, and conceptual capacity". Even so, Shoemaker's view is committed to *State Dependence* because the first-order psychological state and the "second-order belief" have the same core realization and are therefore not distinct existences, which is also why, assuming a suitable degree of rationality and such, having an accurate second-order belief is "automatic".

The first difficulty faced by a *State Dependence* view like Shoemaker's centers on fallibility. Oftentimes we are mistaken about our own psychological states. Since everyone these days recognizes the existence of these kinds of errors, a proponent of *State Dependence* must restrict the scope of his thesis in some way to make room for exceptions.[100] Shoemaker, for instance, restricts the dependency relation to cases where we have "a certain degree of rationality, intelligence, and conceptual capacity". Only then does having a psychological state entail knowing that one has it (only then do the two states have the same "total realizations"). But why is that? What is it about a given degree of rationality intelligence or conceptual capacity that explains why psychological states are self-intimating? Whatever this "certain degree" is, it

[100] This is why Shoemaker understands *State Dependence* as a generic.

must be quite minimal. Small children presumably have first-person access to their psychological lives without having a very sophisticated degree of intelligence or conceptual mastery. But it is hard to understand why a single brain state will always realize, for example, a belief that it is raining but, will only realize the knowledge that one believes it is raining when a degree of rational intelligence is added. How does a "certain degree of rationality, intelligence, and conceptual capacity" change the properties of the brain state? Shoemaker owes us an explanation for why these conditions in particular determine whether or not a person knows about her own psychological states. The problem, I think, is that if we accept *State Dependence*, we *cannot* provide a plausible explanation.

The difficulty stems from the fact that, with one minor exception, there is no limit to the kinds of psychological states about which we make mistakes.[101] Our fallibility is universal; it ranges over any psychological state we might have. For any type of psychological state, it is always possible that another person knows that I have it but I believe otherwise. On Shoemaker's view, this happens because I lack a certain degree of intelligence or rationality. But can that really be true? It seems easy to think of cases where ordinary rational human beings have a desire or a belief but do not know about. At times, we make a simple mistake about what we want for dinner tonight or what we believe about a difficult philosophical issue. When we do, we might have a friend around who corrects our mistake, but it does not seem that we are being especially irrational or that we suddenly lack a level of intelligence.[102]

Rationality, intelligence, and conceptual capacities are relatively stable character traits. But, on Shoemaker's view, it looks as if they must change frequently and dramatically in order to accommodate changes in the epistemic access we have to our psychological states. Suppose that, while focused on writing this dissertation, I am unaware of my desire to listen to country music. If Shoemaker were right, I would lack "certain degree of rationality, intelligence, and conceptual capacity"; otherwise the underlying brain state would realize the knowledge of my desire. But this does not seem to be a very good explanation of my lack of awareness. My lack of awareness of this desire is normally consistent with my maintaining intelligence, rationality and conceptual capacities. What is stranger, is that on Shoemaker's view it seems that I would acquire a "certain degree of rationality, intelligence, and conceptual capacity" when I learn about my desire to listen to country music. It might happen that my roommate plays a country song and I become aware of my desire to hear precisely that kind of music. In this type of case, I gain knowledge of my own psychological life. But do I really also gain rationality, intelligence or conceptual capacities?

---

[101] The exceptions I have in mind are beliefs that are self-verifying in the sense that they involve what Burge calls cogito-like judgments--beliefs like, "I believe that I am thinking". See Burge (1988) and (1996).

[102] For more detailed examples of this kind of ignorance, see the following chapter.

The basic problem with a view like Shoemaker's is not that it makes fallibility impossible but that it makes it unintelligible. For a *State Dependence* view, there is no plausible theory of error available that would account for instances of ignorance or error. One would need to explain why states that are normally ontologically interdependent were not. But the question is how contingent features of a person, such as their degree of intelligence, could affect an ontological dependency. How *could* contingent conditions break that strong of a relation?

Consider an analogy. Suppose P implies Q and I believe both that P and that P implies Q. In favorable conditions I will believe that Q. If I do not, others might say that I am irrational and that is a way of pointing out that conditions are unfavorable. Alternately, I may be quite good at reasoning and not believe that Q because I am distracted by hunger or concentrating on writing a philosophy essay. The point is that, were I to not believe that Q, there would be various explanations available for what makes this situation unfavorable, which *ipso facto* explain my mistake. Sighting *why* conditions are unfavorable just amounts to accounting for my error or ignorance. This is because the logical connection between P and Q functions as the epistemic basis for proper reasoning to the belief that Q. Contingent facts about me can alter whether I believe that Q because of a gap between this epistemic basis and my actual psychological processes, between the beliefs I actually have and the ones I ought to have. We can make sense of a person failing to grasp an epistemic ground or failing to reason appropriately because we can make sense of someone failing psychologically to live up to epistemic norms. But the actual transition from believing that P and that P implies Q, to believing that Q is *not* a logical one. Rather, the first two psychological states cause the third one. So, even though there is a strong connection between the contents of my beliefs, my actual reasoning rests on contingent causal connections that might fail to hold in unfavorable conditions. This gap between these causal relations and the stronger logical relations allows for a plausible theory of error but it also breaks the analogy with *State Dependence*.

The core of *State Dependence* is that *no* contingent causal relation connects my psychological state with my state of knowledge of it. This presents a problem because causal relations allow for clear explanations of why we sometimes are mistaken. As was just illustrated, we are normally fallible because of a gap between the epistemic ground for our beliefs and our own psychological processes. We make mistakes because psychologically or sub-personally things are not working properly. A causal failure might derail a reliable tracking mechanism or, due to a lack of attention, we may fail to grasp the justification or evidence for what we should believe. The problem facing a *State Dependence* is that it eliminates the gap between our psychological processes and the epistemic basis of our reasoning that usually accounts for errors. Because there is no dynamic transition from my psychological state to my knowledge of it, no process can fail to meet epistemic standards. This is the reason *State Dependence* makes our epistemic failures inexplicable. The most

plausible theory of error will appeal to a transition from our psychological states to our knowledge of them, a minimally causal transition that could, on occasion, fail.[103]

Nevertheless, it remains true that we are not usually wrong about what we think or feel; most of the time we get the existence and the character of our psychological states just right and we usually do so in a special first-personal way. But an equally important truth is that a person may be wrong about the existence or character of any one of her psychological states. I have tried to show that we can accept both of these truths about what it is like to for us to have psychological states without giving up the simple thought embodied in the *Distinct Existence Thesis*.

V

Much of what I have said in the preceding sections bears directly on to the issue of first-person *epistemic* authority. It will therefore be helpful to offer some specific remarks on that subject. I have argued that *Capacity Dependence* is true because having a special kind of first-person access is grounded in a person's cognitive agency. This makes it very tempting to think that each of us thereby enjoys an epistemic privilege over others when it comes to knowledge of our own psychological states. Three features of first-person access can easily lead to this temptation. First, other people never have first-person access to my psychological states. So, even if I sometimes lack it, I typically have a special non-evidential way of knowing about my own beliefs, desires, thoughts and feelings that others do not. As we saw earlier, in order for another person to know about my psychological life, she needs to observe my behavior; but this is precisely what I do not seem to need in order to gain knowledge in the distinctively first-personal way. To many philosophers, this fundamental asymmetry between the first-personal and the third-personal ways of knowing shows that first-person access is epistemically privileged.

Secondly, *Capacity Dependence* is a modal claim. Any person who has psychological states in the ordinary first-personal way necessarily has a distinctive capacity to know about them without evidence or observation. But other people do not necessarily have any capacity for epistemic access to my psychological states. Suppose that I were exiled to some far off corner of the Moon. I would, in that case,

---

[103] cf. Boghossian (1989): "I know of no convincing alternative to the following style of explanation: the difference between getting it right and failing to do so (either through ignorance or through error) is the difference between being in an epistemically favorable position with respect to the subject matter in question--being in a position to garner the relevant evidence--and not." (pg. 76) Although Boghossian rightly notices that this is a problem for views like Shoemaker's, he mistakenly believes the only alternative explanations of self-knowledge are either evidential or inferential. The view I have proposed in this chapter is neither.

continue to have psychological states but no one else would have epistemic access to them. The fact that other people have ways of knowing about my psychological life seems to be completely accidental. This modal asymmetry between the two ways of knowing can also make it seem that first-person access is epistemically privileged.

Finally, I have suggested that, because a person must function as a cognitive agent, she is entitled to know about her own psychological states *a priori*. In strictly epistemic terms, our having first-person access just is our having a special *a priori* way of knowing about our own psychological states. In contrast to this, the way other persons know about our psychological states does not seem to be *a priori*. So, someone who accepts my account of the epistemology of first-person access may conclude that it is epistemically privileged in virtue of being *a priori*.

The idea that ordinary first-person access is epistemically privileged is, I think, very widely held. Nevertheless, it is important to see that it is quite a substantial assumption. The basic asymmetry between first and third-personal ways of knowing does not entail that either one is epistemically superior. This asymmetry is recognized, as we saw earlier, by the fact that first-person access does not involve evidence or observation. But, this is a *purely* negative description of first-person access and from it one cannot legitimately infer that self-ascriptions of psychological states are epistemically privileged. As Davidson pointed out, "claims that are not based on evidence do not in general carry more authority than claims that are based on evidence, nor are they more apt to be correct."[104] Thus, our having a distinctive kind of first-person access does not make us into epistemic authorities on our own psychological states.

To see the point more clearly, consider the following analogy. Sometimes when I smile, I know that I am smiling. This is not always true. At times, I can be so wrapped up in what I am doing that I fail to notice that I am smiling. It even happens that someone else can point out to me that I am smiling and I learn that I am through testimony. Regardless, it is quite plausible to think that I often know that I am smiling without evidence or observation. I do not normally have to look in the mirror to know I am smiling but know it simply in virtue of smiling. This should be sufficient to establish and epistemic asymmetry between ways of knowing smiles. In my own case, I often know them without evidence or observation but you can only know I am smiling by observing my face. Does this make my knowledge of my own smiles epistemically privileged? Is my knowledge somehow less fallible, less subject to correction, or more trustworthy? It does not seem that way to me. Rather, it seems that you could be in as good of an epistemic position to know that I am smiling, even though your way of knowing would remain fundamentally different from my own. In order for this difference to ground some epistemic advantage for either of us, much more about the way we know smiles would need to be explained.

---

[104] (1984) pg. 5. cf. Davidson (1987).

Similarly, a far more substantial explanation would be needed to warrant the conclusion that first-person access is epistemically privileged. Such an explanation would require us to supplement the simple characterization of the phenomenon with a more positive one. It would also rest on significant assumptions about the nature of third-personal forms of access to psychological states. For instance, if one thought that another person's knowledge of my psychological states was inferential, that would directly influence how one viewed the epistemic asymmetry between the first and third-person. Someone who wishes to vindicate an epistemic privilege for first-person access will therefore need to go well beyond the most basic features of the phenomenon.

Let's turn to the two more specific reasons a philosopher might be tempted to think first-person access is epistemically privileged. *Capacity Dependence* states a necessary truth about a person with psychological states. It might be thought that because having a first-personal way of knowing one's own psychological states is a *necessary* property of persons it is epistemically privileged over contingent third-personal ways. But the claim that first-person access is epistemically privileged is a claim about the *actual* world. Even if having a capacity for this kind of access is a necessary property of persons, this alone does not imply anything about the epistemic status of knowledge arrived at in this manner. It is possible that third-personal ways of knowing are less fallible, less dubitable, less subject to error and more trustworthy. This may very well be how things are in the actual world.[105] So, even though there is some possible world where first-person access does enjoy a clear epistemic advantage, it may not be our world. One must carefully examine actual ways of knowing in more detail in order to determine whether one way was epistemically privileged.

I have also argued that first-person epistemic access is a necessary consequence of a cognitive agent's having psychological states. If that is correct, cognitive agents have an *a priori* entitlement to know about their own psychological states. It might be thought that because this first-personal way of knowing is *a priori* it is epistemically privileged. But, again, this inference is only valid if we make some substantial philosophical assumptions. One such assumption, which is derived from Kant, is often held. Kant believed that we could have *a priori* knowledge of only necessary truths.[106] But there is no good reason to accept this point. Even for Kant, the concept of the *a priori* principally makes an epistemological distinction between knowledge justified by experience and knowledge justified independently of experience. We can accept this distinction, however, without following Kant in thinking that only necessary truths can be known independently of experience. Another widespread assumption is that another person's knowledge of my psychological states is always empirically based, or, as Kant would say, *a posteriori*.

---

[105] In Chapter 5, I shall argue that this is indeed how things are in the actual world.
[106] Kant (1781), B3-B5.

But we need not hold this view either. Some philosophers, notably Tyler Burge, have argued that third-party knowledge of my psychological states is also warranted *a priori.*[107] Burge's view is controversial and I am not endorsing it here, but if he is right, then the first-personal and third-personal ways of knowing are epistemically on par even though the explanation of why each is an *a priori* way of knowing would differ dramatically.

Someone interested in vindicating the notion that first-person access is epistemically privileged will probably view these remarks as a challenge. Starting with the recognizable epistemic asymmetry between first and third-personal ways of knowing psychological states, a philosopher could attempt to develop a positive account along the lines that I have indicated. The history of philosophy already contains a variety of such accounts, but so far none have been entirely satisfactory. One may, nevertheless, pursue this approach to understanding first-person access but that means require a far more elaborate theory than the one I have offered. I favor taking a more modest approach. I believe one can explain how first-person access is epistemically unique without having to vindicate the idea that it is epistemically privileged. This is what I have attempted to do in this chapter. The modest approach admittedly cannot *explain* why first-person access is epistemically privileged. But what if it is not? What if our way of knowing about our own minds is in fact no better than the ways available to our friends or family? In that case, modesty would be a clear theoretical virtue.

.

[107] Burge (1993).

CHAPTER 5
ILLUSION, ERROR, AND CONSCIOUSNESS

For most of the facts, no person is guaranteed to be in a better epistemic position to know about them than anyone else. Although I may in fact be in the best position to know about the temperature inside of my office, this is not always the case. Sometimes someone else might be in the office and, in that case, he would be in the better epistemic position. Or, we may both occupy the office at the same time and share the same type of epistemic point of view on the temperature. And it is also possible that I never entered the office and so was never in a good epistemic position to know about its temperature. Some of us happen to be in a good position to know some of the facts, while others are in a better position to know others. Thus, whether or not someone's way of knowing something enjoys an epistemic privileged over other ways is typically a contingent matter.

For psychological facts, however, things can appear to be quite different. To many people, it seems reasonable to think that each person is guaranteed to be in an especially good epistemic position to know about her own psychological states. This is largely because each of us has a distinctive first-personal way of knowing about our own psychological states that is never available to anyone else.[108] In order for some other person to know about my psychological states are, they must rely on perception in some way. But I do not normally use perception to find out them. My own way of knowing about my psychological states is therefore recognizably different. Noticing this epistemic asymmetry between ways of knowing can naturally lead to the thought that the first-personal way is epistmically privileged, that each of us not only has a distinctive way of knowing about our own psychological states, but also a much better way of knowing about them.

These kinds of considerations have lead many philosophers to hold the following thesis:

Epistemic Privilege: For any psychological state M, the first-personal way of knowing M is epistemically more reliable than other ways of knowing M.

If you think *Epistemic Privilege* is true, then our ordinary first-personal way of knowing about psychological states is privileged not in the strong sense of infallibility or

[108] The first-personal way of knowing is often called introspection. Because I think the term "introspection" suggests a specific understanding of first-personal knowledge that is unhelpful, I avoid using the term. But, if one takes it to be a neutral concept, "introspection" is roughly synonymous with the first-personal way of knowing I am interested in.

incorrigibility but in the modest sense of being less likely to be mistaken.[109] No other person can know about our psychological states in this first-personal way. So, if third-personal ways are more likely to generate instances of error or illusion, then our ordinary first-personal way of knowing would have a superior epistemic status.

Two important points about *Epistemic Privilege* should be made immediately. First, *Epistemic Privilege* is compatible with the ontological independence of a person's psychological states from her states of knowledge of them. If we assume that psychological states are distinct existences from any states of awareness, this in no way undermines *Epistemic Privilege.*[110] The first-personal way of knowing may, nevertheless, be a privileged way of knowing even though psychological states are ontologically independent from our knowledge of them. Second, we should not infer that *Epistemic Privilege* is true simply on the basis of a clear epistemic asymmetry between first-and third-personal ways of knowing. The truth of *Epistemic Privilege* would require a more substantial account of this asymmetry, an account of why the first-personal way of knowing psychological states is in fact superior and not simply unique. The fact that each of us has a distinctive first-personal way of knowing does not entail that it or any other way of knowing is epistemically privileged.[111]

For this reason, the account I offered of first-person access in the previous chapter does not imply that *Epistemic Privilege* is true. In that chapter, I argued that we each have an a priori entitlement to knowing about our own psychological states in a distinctively first-personal way. But, this warrant can only explain why our psychological self-ascriptions are knowledge, not how a person actually comes to know about the existence of any token psychological state. In order for a person to gain knowledge about her own psychology, there must also be causal relations

---

[109] The point is explicitly made, thought not fully endorsed, by Shoemaker (1994): "Our minds are so constituted, or our brains are so wired, that for a wide range of mental states, one's being in a certain mental state produces in one, under certain conditions, the belief that one is in that mental state. This is what our introspective access to our mental states consists in….The beliefs thus produced will count as knowledge, not because of the quantity or quality of the evidence on which they are based (for they are based on no evidence) , but because of the reliability of the mechanism by which they are produced." (pg. 268).

[110] If we do not assume this, then we might believe that an individual's way of knowing her own psychological states is epistemically privileged because it is ontologocially interdependent with the existence of her own psychological states. This metaphysical view was discussed in Chapter 4. Thus, for the remainder of this chapter, I shall assume that the *Distinct Existence Thesis* is true and focus on the asymmetry in ways of knowing.

[111] This inference is not infrequently made. As I pointed out in Chapter 2, it is an inference Moran makes in his account of first-person authority (Moran 2001 and 2004).

connecting her actual psychological states to her awareness of them. Even if these relations, as I have argued, do not determine our epistemic entitlement--we are not entitled to make knowledgeable self-ascriptions of psychological states *because* of the reliability of any underlying causal processes--they are nevertheless necessary conditions for making knowledgeable self-ascriptions.[112] It might be that the processes realizing our first-personal way of knowing are more reliable that those that realize perception, but this is an open question. The account given in the previous chapter is neutral on whether the first-personal way of knowing is epistemically privileged over third-personal ways of knowing.[113]

Nevertheless, many people have a strong conviction that *Epistemic Privilege* is true, or at least that it is true for our most basic types of psychological states. Oftentimes they hold onto this conviction in the face of evidence that indicates otherwise. Recently, a great deal of empirical work in psychology indicates that *Epistemic Privilege* is in fact not true. Psychologists have recently been offering up numerous cases that point toward the conclusion that, in fact, our ordinary way of knowing ourselves is no better than the ways other people do. In this chapter, I shall present a sample of this research and show how it indicates that *Epistemic Privilege* is false. But many philosophers tend to dismiss the relevance of this research to *Epistemic Privilege*. I think this is because they do not believe its truth is an empirical matter. Rather, it seems to me that relevant empirical work is ignored by those who think that, as a matter of principle, our first-personal way of knowing is less liable to error than any perceptual way of knowing. Thus, before I present the empirical research, I shall first clear up some common misconceptions about the nature of first-person access. More specifically, I shall argue that its nature does not rule out the possibility of illusions and other kinds of errors analogous to those that normally occur in perception.

Whether or not *Epistemic Privilege* is true therefore seems to depend on whether our actual first-personal way of knowing is in fact epistemically privileged over other ways of acquiring knowledge of a person's psychological states. If it is more reliable, less subject to illusions, or less liable to generate errors than third-personal ways of knowing, *Epistemic Privilege* will be true. Rather than taking this truth for grated, we should instead investigate the actual phenomenon. In the main part of this chapter, I shall draw on a sample of recent empirical work to argue that *Epistemic*

---

[112] cf. Peacocke (1999): "But the philosophical lesson is that although there are many deep respects in which self-knowledge cannot be assimilated to perceptual knowledge, there is, even in consciously based self-ascription, reliance on a network of causal relations whose obtaining is by no means necessary." (pg. 245)

[113] This is especially true if one thinks that third-personal ways of knowing psychological states also enjoy an a priori entitlement. See Burge (1993).

Privilege is not true.[114] My argument will be inductive, so it will probably not convince everyone. But although one may continue to adhere to the *Epistemic Privilege* I hope to show that this amounts to an act of faith rather than a plausible hypothesis. Although each person has a special way of knowing about her own psychological states, what makes it special is not that it is an epistemically privileged way of knowing.

I

One may be inclined to believe *Epistemic Privilege* because our first-personal way of knowing seems to be fundamentally different from more third-personal ways of knowing, which are essentially perceptual. This is, for example, what Akeel Bilgrami has in mind when he contrasts the two:

> Knowledge of the world and others paradigmatically involves looking, seeing, hearing things in the world, including what others have to do and say--in general, a testimony of the senses. (Even the testimony of others that the world is thus and so involves hearing what they have to say.) Testimony of the senses, or any inner cognitive analogue of the senses, is precisely what is *not* involved in the ordinary and paradigmatic cases of self-knowledge. (2006, p.p. 89-90)

One difference alluded to in this passage, and mentioned often, is that there is no sensory organ needed to acquire knowledge of one's own psychological states. We do not perceive them in anything like the way we use our eyes or ears to perceive features of our surrounding environment. Even philosophers who explicitly favor a perceptual model for understanding first-personal knowledge do not think that there is some kind of organ of inner sense.[115]

A second difference exists if we think that perceptual knowledge is mediated, that objects in the world first cause sensory impressions or appearances that in turn cause perceptual knowledge; that is, if we think that perception essentially involves

---

[114] This will also mean that we do not need to account for an epistemic dimension to first-person authority. If we are not, in fact, in the best position to know about our own psychological states, we do not need to explain how we are authorities in virtue of being in such a position. Thus, the argument against *Epistemic Privilege* further supports the account of authority in agential terms I offered in Chapter 2.

[115] Armstrong (1968) concedes that there is no introspective sense organ but then goes on to argue that proprioception also lacks an organ of sense. Thus, there is a type of perceptual knowledge, knowledge of one's own bodily position, acquired without the use of an organ.

sense-impressions. On this line of thinking, we would, for example, know that the table is brown because we are immediately confronted with a sense-impression of the table being brown which leads us to believe that the table is brown. There is no analogous intermediary sense-impression in cases of first-personal knowledge. No one thinks that we are aware of our own beliefs or desires because we are aware of an *appearance* of our beliefs or desires.

Whether perceptual knowledge of the table is acquired through an intermediary is a puzzling issue that I will not get into. Many philosophers these days endorse the view that it is not mediated because we have direct access to the table and to its color properties.[116] Yet, even if this is true, there is a lingering difference between perceptual knowledge and first-personal knowledge that the sense-impression view helps to bring out--the possibility of illusion is basic to our understanding of perceptual knowledge. Whatever causal processes realize perceptual knowledge can, when malfunctioning, generate a false belief that is based on some kind of illusion or misperception. Thus, although I currently know that the table is brown, I might have had an illusory experience of the table being brown and, in that case, I would believe falsely that the table is brown because of the illusion. This kind of error is sometimes generated by a causal failure in the perceptual system, a failure that produces an illusory sense-experience. But, even when one's perceptual system is working perfectly well, as Tyler Burge notes "one's perceptual states could in individual instances fail to be veridical."[117] If I have an illusory experience of the table being brown, I will falsely believe that it is brown, even though I am behaving as a responsible believer. There is not much one can do to *not* believe a convincing illusory experience. By contrast, even though we are sometimes mistaken about our own psychological states, it does not seem that we can have an illusory experience of

---

[116] Cassam (2009) compares this understanding of perceptual knowledge with self-knowledge. McDowell (1994) makes the case for perception being direct awareness of the world.

[117] (1988), pg. 120. This, again, is what Burge (1988) calls a "brute error". A brute error, in Burge's sense, occurs when the subject is functioning perfectly well cognitively, rationally, and perceptually. The idea is that although everything can be working well with me, *qua* believer, the world might simply be set up in such a way to lead to misleading perceptual states, as, for instance, when a straight stick is placed in water. But, one reason we might think brute error is impossible in the case of first-personal knowledge is that *everything* in this domain is going on inside of the subject. The state that I know about is my own psychological state and so even if it is presenting itself in an illusory manner, like a stick in water, it seems reasonable to say my cognition is generating the error. So if we take any error about psychological states to be a problem with cognition, brute errors are indeed impossible, but if we instead take a narrower understanding of "brute error", it is unclear to me that they are impossible.

them. It does not seem that we could simply "fail" to have veridical awareness of our own beliefs, desires, and feelings in the same way we can have it of objects in the external world.

The necessity of some kind of sensory organ and the possibility of illusions are both part of our concept of what it is for a way of knowing to be perceptual. This is partly why our first-personal way of knowing is thought to lack both of these features; it is not thought of as a kind of perception.[118] Many people understand these as fundamental differences in the respective natures of perceptual and first-personal knowledge that show *Epistemic Privilege* is true.[119] One idea would be that because first-personal knowledge never requires sensory organs, the causal processes it is realized by are less likely to fail. Alternatively, one might infer that these processes are necessarily more reliable because it is impossible to have an illusory experience of one's own psychological states. There is nothing analogous to having a perceptual state that falsely represents the table as being brown, no psychological sense-impressions or appearances upon which a person might base a false belief about her own psychological life. If either of these two inferences is credible, then a first-personal way of knowing would seem to be epistemically privileged over any way of knowing that involved perception.

II

The first inference is based on a premise about the absence of sensory organs. As it stands, the inference is problematic. Nothing about the mere presence of a sensory organ entails that a causal process involving it is more or less reliable. The fact that a transmission of information must proceed via organs does not obviously entail anything about the process. Suppose, as is probably the case, that any way of acquiring perceptual knowledge requires a more elaborate causal process than any way of acquiring first-personal knowledge; relevant information must always travel along a sequence of relations at least one of which involves organs. If any part of this process breaks, the information will not ultimately generate knowledge, instead, we will have an instance error or ignorance. Does this mean that the overall process is epistemically inferior?

In order to sensibly answer this question, we would have to know whether our actual first-personal way of knowing is in fact more reliable. If two different ways of knowing are equally reliable, then the fact that one was essentially composed

---

[118] This is even true for those who favor a "perceptual" model of understanding our first-personal knowledge. Although these philosophers model it on perception, they also believe it is fundamentally different from perceptual ways of knowing. See Armstrong (1968).

[119] Other distinctions can be found in Shoemaker (1994).

of more parts, some of which were organs, would not increase its chance of error.[120] This line of reasoning lends no *prima facie* support to *Epistemic Privilege.*

This brings me to the second inference. It seems compelling because many people think that illusions about one's own psychological states are impossible. For some types of psychological states, I am inclined to think this is true. Burge, for instance, argues that misperception and illusion are impossible for what he calls "cogito-like judgments" or thoughts that are "self-referential and self-verifying."[121] These include thoughts like "I am thinking about writing this paper," or "I am judging that this coffee is hot." For thoughts like these, there is no gap between the thought known and one's knowledge; there is no room for anything illusory to creep in. Nevertheless, we make many kinds of psychological self-ascriptions and it seems to me that in other cases illusions are possible.

Strictly speaking, any illusions about psychological states will not be perceptual illusions. But understanding the basic perceptual case is nevertheless helpful. Take as an example of a perceptual illusion believing wrongly that the table is brown because strange lighting conditions cause us to experience the table as brown. Such a belief is not based on direct access to the table and its real properties, but on how things seem. It seems like we are seeing a brown table and that is why we believe we are. There are various accounts of what this "seeming" amounts to, but a simple picture of perceptual illusion need not get into them.[122] According to this simple picture, when I am mistaken because of an illusion, two things happen:

False Belief: I believe that P.

Illusory Basis: My belief that P is based directly on how things seem to me.

In order for the *Illusory Basis* to be illusory, how things seem to me will often be equivalent to P, but this is not necessarily the case. When suffering from a perceptual illusion, I might believe that the table is not blue because I to me it seems brown. But I might also believe that the table is not blue because, to me, it seems that I am not seeing a blue table. The way things seem does not always determine what I believe. If I am aware that an illusion is an illusion, how things seem will not cause me to believe anything. Nevertheless, the simple picture of illusion is applicable to

---

[120] The presence of organs would probably help us explain instances of error in the actual world. But that is not enough to conclude *Epistemic Privilege* is true.

[121] Burge (1988) and (1996).

[122] One can have an illusory experience without having a false belief. So, the account I am giving here is not meant to be a fully account of illusion or illusory experience. Rather, I only want to focus on cases where illusion undermines one's epistemic status, where the illusory experience generates unreliable or false beliefs. These are the only cases where the epistemic credentials of perceptual knowledge are at issue.

cases where I wrongly believe that P because of some *Illusory Basis;* cases where I base my belief that P on how things seem when things are unfortunately not how they seem.[123] This simple picture of illusion would be applicable to first-personal ways of knowing if one could have a *False Belief* on an *Illusory Basis*, if a person could falsely believe something about her own psychological state because of how things seemed to her. The problem with applying the simple picture is that we have come to believe that in the psychological realm things are always exactly as they seem; we believe that we could never find anything that could function as an *Illusory Basis*.

Although this is a widely accepted idea, I think it is false. People are able have a *False Belief* about their own desires, moods, and feelings because of how things seem to them. Sometimes when I feel tired, it seems to me that I am sad and I therefore believe that I am. This is a case where how things seem to me is an *Illusory Basis* for a *False Belief*. I find it hard to believe others do not experience similar things. Nevertheless, there is also a particularly vivid and well documented example that illustrates how a person can have a false belief about one of her psychological states because of a kind of illusory experience. It occurs when schizophrenic people experience something called "thought insertion".

A schizophrenic person experiencing thought insertion usually reports having thoughts inserted into her mind. Here are some examples:

> Thoughts come into my head like 'Kill god'. It's just like my mind working, but it isn't. They come from this chap, Chris. They're his thoughts. (Frith, (1992))

> Sometimes it seemed to be her own thought '. . . but I don't get the feeling that it is.' She said her 'own thought might say the same thing . . . But the feeling it isn't the same . . . the feeling is that it is somebody else's . . . (Hoerl, 2001)

In each of these reports, the person with schizophrenia attempts to characterize what his or her experience is like, what things are like from her own point of view. These reports are truly incredible. How could anyone think that thoughts woven into their own psychological life belonged to someone else? What does it even mean to say that "it's just like my mind working" but then disavow that it is? The audacity of these reports might tempt us to treat them as nonsense. But, despite its seeming

[123] I am intentionally ignoring many questions about this basing relation. In particular, whether it needs to be something the subject is aware of or not. For the purposes of this chapter, I don't think it matters. The simple picture is that one lacks perceptual knowledge, when under the guise of an illusion, because one's false belief is based on how things seem.

incoherence, the idea that a person might have thoughts inserted into her mind seems to be something we can imagine. There is no straightforward contradiction to the idea that I am directly aware of thoughts or other mental events that belong to someone else and a part of the reason thought insertion is so intriguing is that what the schizophrenic person reports it is not obviously impossible.

One strange aspect of thought-insertion is the odd beliefs schizophrenic people form to explain the origins of inserted thoughts. Thoughts are universally described as being pushed into the mind by another agent, often in incredibly implausible ways.[124] But, in addition to strange explanations, according schizophrenic people, it seems to a schizophrenic person that thoughts are fundamentally alien and are being pushed into their head. This is how things are from her point of view. This is importantly also how psychiatrists understand the phenomenon:

> Patients report that they feel the thoughts which occur in their heads as not actually their own. They are not experienced as thoughts communicated to them…but it is as if another's thoughts have been engendered or inserted in them. (Cahill and Frith (1996), pg. 278)

It is incredibly challenging to understand precisely what these people are experiencing, what things are like for them. Although it might at first seem that we understand reports of thought insertion, it is difficult to conceptualize them coherently. Perhaps for this reason, psychiatrists struggle with understanding why people experience thoughts as alien or inserted. One recent hypothesis offered by Bortolotti and Broome proposes that "the 'inserted' thoughts can be accessed directly, but are neither acknowledged as one's own nor endorsed on the basis of reasons." (2009, pg. 206) Bortolotti and Broome contend that we normally have direct access to our thoughts and we acknowledge them as our own, as thoughts belonging to us. On this view, a schizophrenic person suffers from thought-insertion because she cannot acknowledge her thoughts to be her own; they float free from what she takes to be her own psychological life.

By describing thought insertion in terms of "acknowledgment" and "endorsement", Bortolotti and Broome make it seem as though a schizophrenic person has a kind of cognitive deficiency. On their view, she has accurate awareness of the content of her own thought, but she fails to acknowledge the thought as belonging to her. But even though a schizophrenic often does fail to acknowledge her thoughts, this cannot be a complete description of the phenomenon. To see this, we need only to imagine a schizophrenic person who does manage to begin acknowledging ownership of her thoughts. Suppose that a reflective schizophrenic person comes to understand that the explanation of thought insertion he has been

[124] Patients have reported, for example, being given thoughts electrically or via raindrops falling on an air conditioner.

giving is plainly ridiculous--no person could really insert thoughts into someone else's mind. This schizophrenic might then come to believe that any psychological state he has direct awareness of necessarily belongs to him. If he does, he could infer that any inserted thought was also one his and, in this way, he could come to *acknowledge* that inserted thoughts are actually his own. But, as long as he continued to experience inserted thoughts as alien there would still be something wrong. Things would still seem to him as if the thoughts he were accessing belonged to someone else.

Something analogous happens with a familiar perceptual illusion. Consider the Muller-Lyer lines. Most philosophers know that these two lines are equal in length. Yet, when we look at the lines, they don't seem to be equal. One line looks to be longer than the other. If we are aware of this fact, we can infer that the two lines are equal and, in this way, *acknowledge* their equivalence, even though that is not how they seem to us. We need not be taken in by every illusory experience. We can form true beliefs despite how things seem. Nevertheless, according to the simple picture of illusion, when we are in the grips of the illusory experience of Muller-Lyer lines, we will believe (*False Belief)* that the lines are unequal because that is how they seem (*Illusory Basis).*

This simple picture of illusion can also be used to accurately characterize the phenomenon of thought insertion. A schizophrenic person has a *False Belief*. He believes that certain thoughts are inserted into his psychological life. Moreover, he believes these thoughts are inserted because that is how things seem to him. Any particular inserted thought seems to be alien, it seems like the product of another mind, or, as Cahill and Frith put it, it seems "as if another's thoughts….have been inserted into them." How things seem functions as the *Illusory Basis* of what a schizophrenic person believes. The schizophrenic is victim to a kind of illusion.

But, is this an accurate characterization of thought insertion? Should we really take the reports of a schizophrenic person at face value when they tell us how things seem to them? They are, after all, extremely ill. The kinds of things a schizophrenic person reports are strange and it might be thought that they present only a marginal case. I am sure that someone will maintain in ordinary cases of first-person access, in cases where one's mind is functioning well, illusions are impossible. Unlike cases of perception, these illusions about psychological states are only possible on the condition of some kind of cognitive defect or pathology.

In one sense, it is true that any illusory experience of a psychological state occurs because of a cognitive deficiency. Perception puts one in touch with objects external to cognitive systems, objects whose existence is independent of cognition. So, perceptual illusions can arise because of something internal or external to one's cognition malfunctions. With thoughts, however, everything is internal; the facts one accesses are internal because they are instances of cognition. Thus, it is true that in this domain all illusions are only possible on the condition of *some* cognitive defect but it also seems trivial. Like perception, it is possible to draw distinctions between

the sources of illusions.[125] Based on the reports of a schizophrenic person, thought insertion does not seem to involve a problem with a person's mode of access to relevant facts. If this is right, the problem that caused thought insertion would still be within the cognitive system, but the first-person way of accessing psychological facts could be functioning fine, just like one's perceptual way of accessing facts functions fine in the Muller-Lyer illusion.

Nevertheless, it is right to note that thought insertion is rare. Everyone seems to fall victim to the Muller-Lyer illusion, but most people do not have illusions of inserted thoughts. My point in discussing the case, however, is only to illustrate that illusory experiences of psychological states are possible, not that they are prevalent. Only if such illusions were impossible could one infer that *Epistemic Privilege* is true on the basis of the *nature* of our first-personal way of knowing. If this way of knowing could not possibly fall victim to illusory experiences, it might be reasonable to disregard empirical research as irrelevant to *Epistemic Privilege* because its truth could be established *a priori*. But, actual cases of illusory experience call this line of reasoning into question. It may be true that our first-personal way of knowing is *less* vulnerable to illusory experience than perceptual ways of knowing and, if so, that alone might indicate *Epistemic Privilege* is true. But this is an empirical point about. The truth of *Epistemic Privilege* cannot be settled without empirical investigation.

III

Existing empirical data undermine *Epistemic Privilege* in at least two ways. First, they show that we are regularly mistaken or ignorant about both the existence and the character of our own psychological states. We self-ascribe beliefs, feelings, motives, personality traits and other attitudes that we do not have and are also completely unaware of many psychological states we do have. Secondly, research shows that when we do know about our own psychological states, third parties usually have as accurate knowledge of those states. It looks as if third-personal ways of knowing a person's psychological states are just as reliable as her own first-personal way of knowing them.

In the remainder of this section, I will discuss only some of this empirical research. Proponents of *Epistemic Privilege* may be tempted to treat the evidence I present as tangential, arguing that it does not speak against the paradigm cases of first-person access. Therefore, in the last section of the chapter, I will address an extremely important objection to my argument: namely that, despite the evidence I present, *Epistemic Privilege* holds true for a special subset of psychological states. But, first I want to consider our initial reaction to the empirical research.

---

[125] For example, consider phantom pain. One way to treat the phenomenon is that it is a case of an illusory pain experience.

III.I

Many social psychologists have recently begun to study what they call implicit attitudes. An implicit attitude is a psychological state outside of a person's awareness that guides her behaviors, judgments, and future decisions. Thus, the existence and character of an implicit attitude is typically revealed only through publicly observable behavior that is equally accessible to everyone. Suppose that I implicitly believe that Oakland is a dangerous city. Since I am not aware of the belief, I would not normally self-ascribe it. In fact, I might say that I believe Oakland is no more dangerous than any other city in California. Nevertheless, I would show my implicit attitude through my actions--consistently avoiding Oakland after dark, taking circuitous routes that bypass the city, refusing to dine at any Oakland restaurants. I believe that Oakland is a dangerous city, but I do so implicitly; unlike my other beliefs, it is not one I am aware of. One defining feature of an implicit attitude is that it is inaccessible in the ordinary first-personal way; it is unavailable to our ordinary way of knowing our psychological states.

If a person can fail to be aware of her implicit attitudes, it seems plausible that she may also be wrong about them. Numerous studies have shown that this is indeed the case. People regularly have an implicit attitude toward an object or group but claim to have the opposite one. Although their self-ascriptions are completely sincere, they are clearly inaccurate. This is shown by the fact that the self-ascribed states do not guide any of the person's subsequent reasoning, judgments, actions or decision making. Unless we naively assume that what a person says about her own psychological states is always correct, this evidence indicates that people are often wrong about what they think, want or feel.

Research on stereotypes offers many specific and clear examples. In a review of the research on stereotypes, Greenwald and Banaji note that "studies suggest that stereotypes are often expressed implicitly in the behavior of persons who explicitly disavow the stereotype." (1995, pg. 14) Similarly, Timothy Wilson concludes from this research that "people can sincerely believe that they are not prejudiced and yet possess negative attitudes at an implicit level." (2002, pg. 133) In one study, for example, subjects explicitly reported having egalitarian beliefs; they said that they were not prejudiced toward racial or ethnic groups. However, when these same individuals were given a basic word association test, they consistently grouped positive character traits with the concept "white" and negative traits with the concept "black".[126] This seems to indicate that the people did not really have the attitudes they reported. Additional research confirms this result; subjects who have prejudiced beliefs implicitly will self-ascribe beliefs with the opposite valence.[127]This

[126] Dovidio et. al., (1986).

[127] See, for instance, Lemm and Banaji (1999), Wilson and Dunn (2004), and Wilson (2002).

phenomenon is not unique to racist attitudes. In one study on gender, Greewald and Banaji (1995) tested whether subjects would associate the concept of "achievement" more with "male" rather than "female". Again, although subjects self-ascribe egalitarian beliefs, the study showed people "associate male gender, more than female gender, with achievement."(1995, pg. 16)

Perhaps we should have expected inaccurate reports when asking people questions about gender or race. What if these inaccuracies are generated by the politically charged nature of the questions? What if these people are inaccurate because they are embarrassed or ashamed? What if errors in self-ascription occur because the attitudes are prejudicial or unpopular? Implicit attitudes are unconscious and no one thought first-person access was a way of knowing about *unconscious* psychological states.

It is important to notice that these people are not simply ignorant of unconscious attitudes. Rather, when they are explicitly invited to access their thoughts and feelings in a first-personal way, they give *wrong* answers. When subjects in the Greenwald and Banaji study, for instance, report having egalitarian beliefs it is because they tried to access their psychological state first-personally. They did not self-ascribe the belief on the basis of behavioral evidence. So, it seems the error is generated by a failure of first-person access.[128] In these cases, the processes that ordinarily realize first-personal knowledge cause the person to falsely self-ascribe egalitarian attitudes. Thus, in these studies, evidence indicates that the causal processes underlying our first-personal way of knowing produce inaccurate beliefs about our own psychological states.[129]

---

[128] See Carruthers (2010) for a different interpretation. He argues that the evidence indicates that there is no such thing as first-person awareness of these kinds of psychological states. Carruthers notably does not discuss prejudicial attitudes and I think this may be because they speak against his "mindreading" view. If the subjects in the experiments I am citing were indeed self-ascribing attitudes on the basis of behavior, we would predict them to self-ascribe the same types of attitudes that third parties do, which they do not. "Mindreading" self-ascriptions would be of prejudicial attitudes, which is what third parties report.

[129] Some psychologists, notably Wilson, Lindsey, and Schooler (2000), argue that the reports of egalitarian attitudes *are* accurate. They suggest that subjects in these cases have, "dual attitudes". They have both the professed egalitarian attitude and the implicit prejudicial attitude. Although I cannot discuss this theory fully here, it strikes me as not the most plausible explanation of the data. Unless we assume that a verbal report is conclusive evidence for the existence of an attitude, a more plausible interpretation of the data is that subjects are mistaken when they report on their attitudes and that their real attitudes are implicit.

III.2

Do we always have first-personal knowledge of our own emotional states? What about our moods or our fleeting feelings? Hardly any contemporary researcher believes that we have a completely accurate grasp of these affective states. These states are oftentimes quite difficult for us to characterize and, even if alarmingly strong emotional reactions are easy to recognize, there seems to be plenty of others that elude us. For example, it seems to be pretty clear that a person can have a mild feeling of anxiety or agitation without being aware of it. When this is the case, people often do not report feeling negative but they also do not avoid self-ascribing psychological states. In fact, it is common to hear someone suffering from a fair amount of anxiety sincerely tell you that they feel fine.[130] In this way, they are making a mistake about their psychological states.

For a mistake like this, consider the following passage from Schwitzgebel (2008):

> My wife mentions that I seem to be angry about being stuck with the dishes again…I deny it. I reflect; I sincerely attempt to discover whether I'm angry--I don't just reflexively defend myself but try to be the good self-psychologist my wife would like me to be--and still I don't see it. I don't think I'm angry. But I'm wrong, of course, as I usually am in such situations: My wife reads my face better than I introspect.[131]

To me, this scenario does not seem far-fetched. In fact it seems relatively common. Sometimes we are so involved in activities and projects that we fail to notice how we are really feeling. When someone asks us how we feel, we self-ascribe something that seems right to us but is not an accurate characterization of any aspect of our psychological life. In this passage, Schwitzgebel is even mistaken about his emotional states after reflecting upon them. He tries to access his psychological states in a first-personal way--he tries to attend to them in order to give an accurate response to his wife--but he comes up with a mistaken self-ascription. As he says, his wife is a more reliable source of information about his emotional states than he is.

---

[130] Haybron (2007) goes so far as to claim that this sort of error makes sense, claiming that "it may be adaptive in may situations to be anxious while *believing* that one is happy, and more generally *presenting* oneself as happy." (pg. 400)

[131] pg. 252. In the remainder of his paper, Schwitzgebel considers a variety of other states of which we may be mistaken. Importantly, he recounts cases in which persons seem to be mistaken about the character of immediate conscious experiences. I will return to these cases in section IV.

Psychological research seems to support the natural idea that people are mistaken about their feelings and moods.[132] An astonishing example of this is that people seem to be mistaken about whether or not they are happy. In response to global happiness surveys, an incredible number of individuals report being happy--up to 94% in one study.[133] Even more striking is that almost no one reports being unhappy. According to these surveys, nearly all humans are either moderately or very happy. Should we trust these self-ascriptions? Should we think that these people have accurate first-personal knowledge of their emotions and that pretty much everyone is happy? Or, instead, should we think that the self-ascriptions of happiness are wrong?

One reasonable way to approach these questions is to consider alternative ways of assessing a person's happiness. Daniel Haybron (2007) considered a handful of relevant studies that tested whether people were happy without relying on their self-ascriptions of happiness. The results were radically different from those in the global surveys. Haybron discusses one study that found that people have "feelings of sadness, fear, anger, or fatigue…34 percent of the time."(2007, pg. 410) He rightly suggests that people experiencing these feelings for over one-third of their day are unlikely to be as happy as the surveys suggest. Haybron concludes that "we should take seriously the possibility that very many people are substantially mistaken about how happy they are."(2007, pg. 411) If he is right and Schwitzgebel is right, other people are often better placed to know about our emotions and feelings.

III.3

One paper is cited more often than any other as evidence for the unreliability of first-person access. After reviewing a large segment of existing empirical research and conducting additional studies of their own, Nisbett and Wilson (1977) provocatively argued that "the accuracy of subjective reports is so poor as to suggest that any introspective access that may exist is not sufficient to produce generally correct or reliable reports." (pg. 233) Any research that could support this conclusion speaks directly against *Epistemic Privilege*. What Nisbett and Wilson found was that in a variety of studies people consistently failed to know the reasons why they responded the way they did in certain situations. The participants in these studies were asked to make choices between alternatives, solve problems, or reflect on their present feelings and, in every case, they seemed to have no idea why they acted, judged, or felt the way they did.

I would like to consider just one of the studies conducted by Nisbett and Wilson on the effect the position of objects has on a person's choice. In this study, subjects were asked to choose the single item out of a set that they thought was of

[132] Wilson (2002) in Chapters 6 and 7 discusses a range of this research.
[133] Inglehart and Klingemann (2000).

the best quality. The set was arranged from left to right on the table in order to test the hypothesis that the position of an item determined the subject's choice. The results confirmed this hypothesis. Subjects overwhelmingly prefer right-most items.[134] Nisbett and Wilson report:

> When asked about the reasons for their choices, no subject ever mentioned spontaneously the position of the article in the array. And, when asked directly about a possible effect of the position of the article, virtually all subjects denied it. (1977, pg. 244)

Not only do the subjects in this study not know about the effect positioning had on their choices, when they were told about it, they denied it having any role. Instead, after accessing their own thoughts in a first-personal way, these people say that their preference for the right-most item was based on another reason. They would claim that they based their choice on the qualities of the available objects. Contrary to what they say, however, the research shows these people are incorrect. As Wilson and Dunn put it in a more recent paper, "people do not have complete access to the actual reasons behind their feelings, attitudes, and judgment and thus generate reasons that are consistent with cultural and personal theories and are accessible to memory."(2004, pg. 17.13)

But, if people are wrong about the reasons behind their feelings or judgments, how accurate can they possibly be about the character of those same feelings or judgments? It seems to me that these studies give us grounds for skepticism about the accuracy of the subjects' psychological self-ascriptions. Krista Lawlor also thinks it makes sense to be skeptical about the psychological self-ascriptions people make in these studies. She notes of these subjects that "while they undertake deliberative questions about what their attitudes are to be, subjects report attitudes that we have reason to think will not in the end guide their behavior."[135] Although this is not universally true, it seems to be precisely what happens in most cases.[136] The self-ascriptions made in the context of the experiment do not accurately reflect the psychological states that ultimately guide the person's future judgments, actions or reasoning. Even a short duration after the experiment, subjects will self-

---

[134] In one case, the right-most item was preferred four to one.

[135] Lawlor (2003), pg. 559. What is curious about Lawlor's reading of these experiments is that she also thinks that subjects in these experiments are correct about their current state of mind and that they therefore have epistemic authority. But if "we have reason to think" that their self-ascriptions will not guide their behavior it seems to me we have reason to think their self-ascriptions are false and therefore not epistemically authoritative.

[136] It is not, for instance, true in all cases of position-effect biases.

ascribe completely different attitudes than those they did during the experiment.[137] Since, subjects do not in fact self-ascribe psychological states that "guide their behavior", their initial self-ascription seems to be mistaken. Commenting on this in a more recent book, Wilson says that people "construct a story about how they feel that is based on reasons that are not entirely trustworthy. The story has the ring of truth to people, but because they have used faulty information…it often misrepresents how they really feel."[138] The research seems to indicate that first-person access is not always the most reliable measure of a person's actual psychological states.

III.4

So far, I have been focusing on research that shows people are often mistaken or unaware of their psychological states. But this same research indicates that when people do know about their own psychological states, others know them just as well.[139] The knowledge other people have requires perception, so, if their reports are as reliably accurate, *Epistemic Privilege* does not seem to be true.

This equivalence in accuracy between self-ascriptions and the reports of others was noticed by Nisbett and Wilson. They discovered in all their studies that third party observers "made predictions that in every case were similar" to the self-ascriptions given by the subject.[140] Whatever people were saying about their own psychological lives, strangers observing them would say the same thing. Nothing about the person's having first-person access helped them give a more accurate report of their psychological states.

Moreover, when the third-party is *not* a complete stranger, as in the Nisbett and Wilson studies, but rather someone familiar with the subject's behaviors, facial expressions, or tones of voice, they often *more* accurate about the person's psychological states. This was already suggested in Schwitzgebel's anecdote. In his anecdote, his wife is a more accurate judge of whether he is irritated or angry. Recent studies confirm this kind of thing to be true in many cases. Recent research studying the impact that close relationships have on knowledge of psychological states indicates that our close friends and family members are at least as reliable at

[137] Evidence for this can be found in Wilson and Schooler (1991), Wilson, Hodges, and LaFleur (1995) and Hodges and Wilson (1994).
[138] Wilson (2002), pg. 169.
[139] Wilson, for example, remarks that "it is remarkable…that a personal advantage over strangers' reports has been difficult to find." (2002)
[140] (1977) pg. 247.

discerning our psychological states and, in many instances, they are much more reliable.[141]

This last point is crucial to notice because, if *Epistemic Privilege* were true, any evidence indicating that other people, whether strangers or close friends, had as accurate knowledge of our own psychological states would have to be an anomaly. If our first-personal way of knowing were epistemically superior to other ways of knowing, other persons would not *normally* be as accurate in assessing our psychological states. Even if they had an epistemic advantage in some cases, these would be rare; the greater reliability of our first-personal way of knowing would be manifest in the vast majority of cases. But, this does not seem to be the case. The kind of situation described by Schwitzgebel seems to occur regularly. Quite often our close friends or our family members are better at gauging our true feelings, especially when our attention is diverted or we are distracted. The fact that people who are close to us are as reliable at assessing our psychological states is the norm, not the exception.

My point in not that we do not sometimes know what we are thinking or feeling. We usually do. And it is not that we do not sometimes know about what we are thinking or feeling when other people do not. That also happens. But my knowing something about my own psychology when other people are ignorant of it isn't enough to conclude *Epistemic Privilege* is true, especially if, they sometimes know things about me that I do not. Given our best empirical and scientific understanding of the processes that realize our first-personal way of knowing, there does not seem to be a reason to think they are more reliable or accurate. I doubt that some further empirical investigation into will reveal something different. Even if *Epistemic Privilege* seems intuitively plausible, it cannot ultimately stand up to empirical research about how we know our own psychological lives.

IV

It is likely that a defender of *Epistemic Privilege* will attempt to restrict its scope to a subset of psychological states. In his more recent book, Wilson has even retracted

---

[141] Colvin and Funder (1991). See also Vazire (2010). Vazire attempts to show that others are consistently more accurate for a range of psychological states, those she calls observable, while a person is more accurate for neuroticisim-related traits. This, however, is only true when others are strangers. When they are close acquaintances, the self's epistemic advantage on neuroticism-related traits also disappears. Although some of her hypotheses are only "partially supported", her work would be especially relevant to the objection considered in section IV, which attempts to isolate a subset of psychological states where first-person access is necessarily more reliable.

his earlier conclusions about epistemic privilege, stating instead that "to the extent that people's responses are caused by the conscious self, they have privileged access to the actual causes of these responses."(2002, pg. 106) Wilson's thought about consciousness is very natural. Cases in which first-person access is impaired, including all those mentioned in the previous section, all involve psychological states that lie outside of consciousness. But, as traditionally conceived, our first-personal way of knowing is a way of coming to know about our *conscious* states. We might grant that others can be as accurate at discerning things about prejudicial beliefs or hidden feelings. But it seems that a person will always have a better way of knowing about her own conscious states than anyone else.

This line of thinking is extremely important to understand because it embodies a nearly axiomatic idea in the history of philosophy. Many philosophers believe that first-personal knowledge is intimately connected to consciousness. To say that a particular psychological state is a conscious state is just to say that it is directly accessible from the first-person point of view and to say that we have first-person access or knowledge of a psychological state is to just to say that we have knowledge of *conscious* psychological state.[142] On this of thinking, a conscious state stands in an epistemically special relationship to our first-personal way of knowing--it is uniquely positioned to be known in this way. No similar relations connect objects in the external world with our perceptual knowledge. One can easily conceive of material objects that we are completely unaware of or ones that are perceptually inaccessible. But, it is thought, one cannot conceive of a conscious psychological state that we are unaware of and one can certainly not conceive of conscious states that are inaccessible. If this is true, *Epistemic Privilege* looks to be true for conscious states.

The nature of conscious states has recently received a tremendous amount of attention philosophers. But, I am doubtful that contemporary discussions have a single thing in mind. For this reason, it is difficult to assess claims that connect conscious states to our first-personal way of knowing. The general thesis seems to be that by having the property of consciousness a psychological state becomes especially well-suited for first-person access; something about the state's being conscious makes it available to be known in this way. This seems to be a metaphysical thesis about the nature of consciousness. It is by having this property of consciousness that a psychological state becomes accessible in a first-personal way. This sort of idea is what Ryle had in mind when he criticized those who thought of conscious states as "self-luminous" entities.[143]

---

[142] David Rosenthal goes so far as to claim that "we have introspective access to mental states only when they are conscious." (1986, pg. 474)

[143] Ryle (1949) Chapter 6.

One way to clarify the general thesis would be to think of the property of consciousness in epistemic terms. This is, for instance, the way David Rosenthal characterizes consciousness:

> A state is conscious if whomever is in it is to some degree aware of being in it in a way that does not rely on inference, as that is ordinarily conceived, or on some sort of sensory input. Conscious states are simply those states to which we have noninferential and nonsensory access. (1986, pg. 464)

According to Rosenthal, a conscious state just is a state that one is aware of. If this is right *Epistemic Privilege* would be true for conscious states but it would be trivially true.[144] A view that defines the property of consciousness in epistemic terms will have as a consequence the fact that a person knows all of her conscious states.

If *Epistemic Privilege* is going to be a substantive thesis about conscious states, the property of consciousness must be understood in terms that do not directly imply that a person has awareness or knowledge of her conscious states. We might therefore choose to adopt David Chalmers definition of a conscious state as one "there is something it is like to be in." (1996, pg. 4) This remark picks up on a very popular idea that being in a consciousness state is being in state with "something it is like". Unlike Rosenthal's definition it does not even partially understand the property of consciousness in epistemic terms.[145] If a conscious state is just a state with something it is like, then it is not trivial to say that the property of consciousness makes psychological state well-suited for being known in a first-personal way; for there is no obvious reason to think states for which there is "something it is like to be in it" are especially well-suited for being accessed in a first-personal way. Just because there is "something it is like" to be in a particular psychological state does not entail that I know what it is in fact like.

This last claim will strike many people as plainly wrong. Some people will think that if there is something it is like for me to be in a particular psychological state, then I know what this is. If I did not know what it was like, then there would not *be* anything it is like *for me*. It may seem *prima facie* incredible that a conscious state, a state that has a kind of subjective character, could even be something that a person is not aware of in a special way. But why is that the case? Why could someone be unaware of a state with a subjective character?

Although it is common to define a conscious state as one with something "it is like" to be in, it does not seem to me that "what it is like" is well understood. Some philosophers think of it as the phenomenal or qualitative properties of a state.

---

[144] For a similar complaint against some theories of consciousness see Sosa (2003).

[145] The idea that consciousness is associated with "what it is like" comes from Nagel (1974), although it is not clear to me, as it seems to be to some, that Nagel intends the phrase to be a definition of consciousness.

But, however we choose to characterize this feature of "what it is like", there is no obvious reason we cannot be unaware of a conscious state we are in. Fred Dretske offers a fairly simple version of a change-blindness experiment to makes this point.[146]

In Dretske's example there are two figures, Alpha and Beta, which have one minor difference--the presence of an additional spot in figure Beta. Someone who looks at each figure will have two qualitatively different experiences, one of Alpha and one of Beta. Since there was a perceived difference between the two figures, observers had a conscious experience of the difference. Yet, as Dretske points out, they are not, or need not, be aware of their experience of the difference; they need not be aware that they are consciously experiencing the difference between Alpha and Beta. Putting this into Chalmers' language, there is something it is like to experience Alpha and something different it is like to experience Beta, making them distinct conscious experiences. There is also something it is like to experience the difference between Alpha and Beta. If consciousness is defined by "what it is like to be in it", each of these is a distinct conscious experience. Yet, as Dretske notices, we are *not* aware of what it is like to experience the difference between Alpha and Beta.

Dretske is not alone in calling attention to the independence of consciousness from our first-personal way of knowing. In a variety of cases like the one Dretske describes, it seems that people can be ignorant of their own conscious states.[147]I suspect this has traditionally seemed implausible because many people have unwittingly assumed a partially epistemic interpretation of consciousness. This epistemic characterization is difficult to avoid and some will surely resist drawing conclusions about consciousness and first-person access from the Dretske's experiments because they think that our consciousness must be connected to our awareness. Change-blindness is contrived and it will be argued by some that, even if we can at times lack awareness of the peripheral conscious states in the change-blindness experiments, we cannot lack first-personal awareness of those states at the very center of our consciousness. These are fundamental conscious states are they are what we must focus on when assessing the truth of *Epistemic Privilege*.

Epistemic Privilege is an empirical hypothesis that ultimately must be evaluated on the basis of empirical evidence. Its plausibility as a hypothesis will therefore depend on how one interprets evidence like that already uncovered by scientists. If this evidence starts to indicate that the hypothesis may be false, defenders can restrict the scope of *Epistemic Privilege* to a subset of psychological states. This is the kind move that prompts the view that *Epistemic Privilege* is only true for *conscious* psychological states. If additional evidence makes that also appear to be false,

---

[146] The point is made by Dretske in his (2004). There is a vast literature on the phenomenon of change blindness. However, I think the simplicity of Dretske's example makes the point especially vivid. For more research see Block (2008) and (2007).

[147] Among others, see Block (2007), (2008) and Schwitzgebel (2008).

defenders can restrict the scope even further to a subset of conscious states, something like *fundamental* or *centrally* conscious states. But making these *ad hoc* modifications to our theories greatly inhibits our understanding of both consciousness and the epistemology of the first-person. If the property of consciousness is not sufficient to explain why we have knowledge of our own psychological states, it is difficult to see what difference fundamentality could make.[148]

Psychological evidence continues to indicate more and more that we do not have any epistemic advantages or privileges with respect to our own psychological lives. I am personally convinced by this evidence that *Epistemic Privilege* is false. In general, we do not have any more accurate knowledge of any type of our own psychological states than others. The first-person point of view does not offer a privileged position for knowing psychological states.[149] Yet, it seems clear that we do have a special way of knowing about our own psychological states, a way that is necessarily unavailable to any other person. It also seems true that we have some authority on our own psychological states. These last two simple observations have been the foundations of the most important accounts of the first-person in the history of philosophy. But, given their traditional epistemic interpretation, they are in direct tension with recent evidence that the first-person point of view lacks an epistemic privilege. Ideally, we should be able to hold onto both the thought that we know about our own psychological states in a distinctive manner and the thought that each of us is an authority on our own psychological lives and we should be able to do this consistently with all current and future scientific research. My goal throughout this entire dissertation has been to illustrate how we can succeed in doing

---

[148] Much more needs to be said about the property of consciousness. Conscious experiences, like pains, after-images, and the like have long been thought to be known in an epistemically privileged way. Understanding how the property of consciousness does or does not affect our first-personal way of knowing is a very complex issue beyond the scope of this dissertation. My own view is that an epistemic gloss on consciousness is behind the assumption that we cannot help but know about our own conscious states. I suspect that once we disentangle qualitative properties of conscious states like pains and tickles from epistemic properties, it will not seem odd to think that a person can be unaware of her own conscious states. I also suspect that a far more precise account of consciousness as "what it is like" will make the surrounding issues tractable. I do not, unfortunately have space here to delve into either of these subjects but they are worthy of far more discussion.

[149] This is not to say that all the questions concerning our first-personal way of knowing are settled. In particular, it seems to me that we know very little about the processes that realize this knowledge. These questions look challenging, especially if we acknowledge that different causal processes might be involved, which is convincingly argued by Prinz(2004)

this. If the first-person is the point of view of a cognitive agent, it need not be epistemically privileged in order to be special.

BIBLIOGRAPHY

Armstrong, David. 1968. *A Materialist Theory of Mind*. London: Routledge and Kegan Paul.

Ayer, A. J. 1963. "Privacy". In his *The Concept of a Person and Other Essays*. New York: Macmillan.

Bar-On, Dorit and Douglas Long. 2001. "Avowals and First-Person Privilege". *Philosophy and Phenomenological Research* (52): 311-335.

Bar-On, Dorit. 2004. *Speaking My Mind*. Oxford: Oxford University Press.

Bilgrami, Akeel. 2006. *Self-Knowledge and Resentment*. Cambridge: Harvard University Press.

— 2010. "Replies to Tom Baldwin and Calvin Normore". *Philosophy and Phenomenological Research* (81) 783-808.

Block, Ned. 2007. "Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience." *Behavioral and Brain Sciences* (30): 481-548.

— 2008. "Consciousness and Cognitive Access". *Proceedings of the Aristotelian Society* (108): 289-317.

Boghossian, Paul. 1989. "Content and Self-Knowledge". In *Externalism and Self-Knowledge*, eds. Peter Ludlow and Norah Martin. Stanford: CSLI Publications.

Bortolotti, Lisa and Matthew Broome. 2009. "A Role for Ownership and Authorship in the Analysis of Thought Insertion". *Phenomenology and the Cognitive Sciences* (8): 205-224.

Boyle, Matthew. 2009. "Two Kinds of Self-Knowledge". *Philosophy and Phenomenological Research* (78): 133-164.

Broughton, Janet. 2008. "Self-Knowledge". In *A Companion to Descartes*, eds. Janet Broughton and John Carriero. Oxford: Blackwell.

Burge, Tyler. 1988. "Individualism and Self-Knowledge". In *Externalism and Self-Knowledge*, eds. Peter Ludlow and Norah Martin. Stanford: CSLI Publications.

— 1993. "Content Presevation". *Philosophical Review* (102): 457-488.

— 1996. "Our Entitlement to Self-Knowledge".In *Externalism and Self-Knowledge*, eds. Peter Ludlow and Norah Martin. Stanford: CSLI Publications.

— 1998. "Reason and the First Person". In *Knowing Our Own Minds*, eds. Crispin Wright, Barry Smith and Cynthia MacDonald. Oxford: Oxford University Press.

Byrne, Alex. 2005. "Introspection" *Philosophical Topics* (33): 79-104.

Cahill, Connie and Christopher Frith. 1996. "A Cognitive Basis for the Signs and Symptoms of Schizophrenia". In *Schizophrenia: A Neuropsychological Perspective*, eds. C. Pantelis, H. E. Nelson, and T. R. E. Barnes. New York: John Wiley and Sons.

Campbell, John. 1999. "Schizophrenia, the Space of Reasons, and Thinking as a Motor Process". *The Monist* (82): 609-625.

Carriero, John. 2009. *Between Two Worlds*: *A Reading of Descartes' Meditations*. Princeton: Princeton University Press.
Carruthers, Peter. 2010. "Introspection: Divided and Partially Eliminated". *Philosophy and Phenomenological Research* (80): 76-111.
Cassam. 2009. "The Basis of Self-Knowledge". *Erkenntnis* (71): 3-18.
Chalmers, David. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
Colvin, C. Randall and David Funder. 1991. "Predicting Personality and Behavior: A Boundary and the Acquaintanceship Affect". *Journal of Personality and Social Psychology* (60): 884-894.
Davidson, Donald. 1982. "Rational Animals". In his *Subjective, Intersubjective ,Objective*. Oxford: Oxford University Press.
— 1984. "First-Person Authority". In his *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press.
— 1987. "Knowing One's Own Mind". In his *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press.
Descartes, Rene. 1641. *Meditations on First Philosophy*. In *The Philosophical Writings of Descartes Volume II,* trans. John Cottingham, Robert Stoothoff and Dugald Murdoch. Cambridge: Cambridge University Press.
Dovidio, John, Nancy Evans and Richard Tyler. 1986. "Racial Stereotypes: The Contents of their Cognitive Representations" *Journal of Experimental Social Psychology* (22): 22-37.
Dretske, Fred. 2004. "Change Blindness". *Philosophical Studies* (120): 1-18.
Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.
Freud, Sigmund. 1911. "Two Concepts of Mental Functioning". In *The Standard Edition of the Complete Psychological Works of Sigmund Freud Volume XI*. trans. James Strachey. London: Hogarth Press.
Fricker, Elizabeth. 1998. "Self-Knowledge: Special Access versus Artefact of Grammar-A Dichotomy Rejected". In *Knowing Our Own Minds*, eds. Crispin Wright, Barry Smith and Cynthia MacDonald. Oxford: Oxford University Press.
Finkelstein, David. 2003. *Expression and the Inner*. Cambridge: Harvard University Press.
Frith, Christopher. 1992. *The Cognitive Neuropsychology of Schizophrenia*. Hillsdale, NJ: Erlbaum.
Greenwald, Anthony and Mahzarin Banaji. 1995. "Implicit Social Cognition: Attitudes, Self- Esteem, and Stereotypes". *Psychological Review* (102): 4-27
Goodman, Nelson. 1954. *Fact, Fiction, and Forecast*. Cambridge: Harvard University Press.
Haybron, Daniel. 2007. "Do We Know How Happy We Are? On Some Limits of Affective Introspection and Recall. *Nous* (41): 394-428.
Heal, Jane. 2001. "On First-Person Autority" *Proceedings of the Aristotelian Society* (102): 1-19.

— 2004. "Moran's Authority and Estrangement". *Philosophy and Phenomenological Research* (69): 427-432.
Heil, John. 1992. "Privileged Access". In *Externalism and Self-Knowledge*, eds. Peter Ludlow and Norah Martin. Stanford: CSLI Publications.
Hieryonymi, Pamela. 2005. "The Wrong Kind of Reason". *Journal of Philosophy* (102): 437-457.
— 2008. "Responsibility for Believing". *Synthese* (161): 357-373.
Hodges, Sara and Timothy Wilson. 1994. "Effects of Analyzing Reasons on Attitude Change: The Moderating Role of Attitude Accessibility". *Social Cognition* (11): 353-366.
Hoerl, Christoph. 2001. "On Thought Insertion". *Philosophy, Psychiatry and Psychology* (8): 189-200.
Inglehart, Ronald and Hans-Dieter Klingemann. 2000. "Genes, Culture, Democracy, and Happiness". In *Culture and Subjective Well-Being,* eds. Ed Diener and Eunkook Suh. Cambridge: Cambridge University Press.
Jackson, Frank. 1982. "Epiphenomenal Qualia". In *Philosophy of Mind: Classical and Contemporary Readings*, ed. David Chalmers. Oxford: Oxford University Press.
Kant, Immanuel. 1781. *Critique of Pure Reason.* trans. Paul Guyer and Allen Wood. Cambridge: Cambridge University Press.
Lawlor, Krista. 2003. "Elusive Reasons: A Problem for First-Person Authority". *Philosophical Psychology* (16) 549-565.
Lemm, Kristi and Mahzarin Banaji. 1999. "Unconscious Attitudes and Beliefs about Men and Women". In *Perceiving and Performing Gender,* eds. U. Paseor and F. Braun. Opladen: Westdutscher Verlag.
Lewis, David. 1983. "What Experience Teaches". In his *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
— 1997. "Finkish Dispositions" In his *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
Locke, John. 1689. *Essays Concerning Human Understanding.* Reprint, ed. P. H. Nidditch. Oxford: Oxford University Press, 1975.
MacFarlane, John. 2011. "What is Assertion?" In *Assertion*, eds. Jessica Brown and Herman Cappelen. Oxford: Oxford University Press.
Martin, C. B. 1994. "Disposition and Conditionals". *Philosophical Quarterly* (44): 1-8.
Martin, Michael. 1998. "An Eye Directed Outward". In *Knowing Our Own Minds*, eds. Crispin Wright, Barry Smith and Cynthia MacDonald. Oxford: Oxford University Press.
McDowell. John. 1982. "Criteria, Defeasibility and Knowledge". In his *Meaning, Knowledge, and Reality*. Cambridge: Harvard University Press.
— 1994. *Mind and World.* Cambridge: Harvard University Press.
Moran, Richard. 2001. *Authority and Estrangement.* Princeton: Princeton University Press.

— 2003. "Response to O' Brien and Shoemaker". *European Journal of Philosophy* (11): 402-419.
— 2004. "Replies to Heal, Reginster, Wilson, and Lear". *Philosophy and Phenomenological Research* (69): 455-472.
— 2005. "Problems of Sincerity". *Proceedings of the Aristotelian Society* (105): 341-361.
Nagel, Thomas. 1974. "What is it like to be a Bat". *Philosophical Review* (74): 435-450.
Nisbett, Richard and Timothy Wilson. 1977. "Telling More than We Can Know: Verbal Reports on Mental Processes". *Psychological Review* (84): 231-259.
Peacocke, Christopher. 1999. *Being Known*. Oxford: Oxford University Press.
— 2008. *Truly Understood*. Oxford: Oxford University Press.
Prinz, Jesse. 2004. "The Fractionation of Introspection". *Journal of Consciousness Studies* (11): 40-57.
Raz, Jospeh. 1997. "The Active and the Passive". *Aristotelian Society Supplementary Volume* (71): 211-228.
Rosenthal, David. 1986. "Two Concepts of Consciousness". In *The Nature of Mind,* ed. David Rosenthal, Oxford: Oxford University Press.
Russell, Bertrand. 1912. *The Problems of Philosophy*. Indianapolis: Hackett Publishing .
Ryle, Gilbert. 1949. *The Concept of Mind*. Chicago: University of Chicago Press.
Scanlon, T. M. 1998. *What We Owe To Each Other*. Cambridge: Harvard University Press.
Schwitzgebel, Eric. 2008. "The Unreliability of Naïve Introspection". *Philosophical Review* (117): 245-273.
Shah, Nishi and J. David Vellemen. 2005. "Doxastic Deliberation". *Philosophical Review* (114): 497-534.
Shah, Nishi. 2006. "A New Argument for Evidentialism". *Philosophical Quarterly* (56): 481-498.
Shoemaker, Sydney. 1988. "On Knowing One's Own Mind". In his *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.
— 1994. "Self-Knowledge and 'Inner Sense'". In his *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.
— 2003. "Moran and Self-Knowledge". *European Journal of Philosophy* (11): 391-401.
Smith, Barry. 1998. "On Knowing One's Own Language". In *Knowing Our Own Minds*, eds. Crispin Wright, Barry Smith and Cynthia MacDonald. Oxford: Oxford University Press.
Sosa, Ernest. 2003. "Consciousness and Self-Knowledge". In *Privileged Access: Philosophical Accounts of Self-Knowledge*, ed. Brie Gertler. Aldershot: Ashgate Publishing.
Squires, Roger. 1969. "Memory Unchained". *Philosophical Review* (78): 178-196.
Stroud, Barry. 2011. *Engagement and Metaphysical Dissatisfaction*. Oxford: Oxford University Press.

Valzire, Simine. 2010. "Who Knows what about a Person? The Self-Other Knowledge Asymmetry (SOKA) Model". *Journal of Personality and Social Psychology* (98): 281-300.
Williams, Bernard. 2002. *Truth and Truthfulness*. Princeton: Princeton University Press.
Wilson, Timothy. 2002. *Strangers to Ourselves*: *Discovering the Adaptive Unconscious*. Cambridge: Harvard University Press.
Wilson, Timothy and Elizabeth Dunn. 2004. "Self-Knowledge: Its Limits, Value, and Potential for Improvement". *Annual Review of Psychology* (55): 493-518.
Wilson, Timothy, Sara Hodges, and Suzanne LaFleur. 1995. "Effects of Introspecting about Reasons: Inferring Attitudes from Accessible Thoughts". *Journal of Personality and Social Psychology* (68): 28-35.
Wilson, Timothy, Samuel Lindsey and Tonya Schooler. 2000. "A Model of Dual Attitudes". *Psychological Review* (107): 101-126.
Wilson, Timothy and Jonathan Schooler. 1991. "Thinking too much: Introspection can Reduce the Quality of Preferences and Decisions". *Journal of Personality and Social Psychology* (60): 181-192.
Wittgenstein, Ludwig. 1958. *Philosophical Investigations*. Oxford: Blackwell.
Wright, Crispin. 1998. "Self-Knowledge: The Wittgensteinian Legacy". In *Knowing Our Own Minds*, eds. Crispin Wright, Barry Smith and Cynthia MacDonald. Oxford: Oxford University Press.