UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

RETHINKING INTROSPECTION:

HOW WE KNOW (AND FAIL TO KNOW) OUR OWN MINDS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

By

JESSE WADE BUTLER
Norman, Oklahoma
2006

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/982

UMI Number: 3206972

Copyright 2006 by
Butler, Jesse Wade

All rights reserved.



UMI Microform 3206972
Copyright 2006 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

RETHINKING INTROSPECTION:
HOW WE KNOW (AND FAIL TO KNOW) OUR OWN MINDS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF PHILOSOPHY

BY

_____
James Hawthorne, Chair

_____
Reinaldo Elugardo

_____
Wayne D. Riggs

_____
Chris Swoyer

_____
Lynn D. Devenport

© Copyright by Jesse Wade Butler 2006
All Rights Reserved

ACKNOWLEDGEMENTS

I would first like to thank my advisor Jim Hawthorne for his guidance and encouragement throughout the development of this dissertation. His feedback has provided an ideal mix of critical reflection and constructive support and has been the basis for numerous improvements to my work. Jim, it has been a privilege to have you as an advisor.

I would like to express my appreciation of my other committee members as well: Ray Elugardo, Wayne Riggs, Chris Swoyer, and Lynn Devenport. These gentlemen have provided valuable instruction, guidance, and insight throughout my work as a graduate student, from my preparatory coursework to the finishing refinements of this dissertation. Having had each of them as an instructor, I can honestly say that they are all outstanding educators. I thank you all for your unique and substantial contributions to my educational development.

I am also deeply grateful for the continual support and inspiration provided by my friends and family. Without the love and encouragement of the many people in my life, this dissertation would simply not have been possible. I would like to give an individual thanks to each of you, but I am fortunate enough to say that that would make this acknowledgement section too long! However, a few special mentions are still in order. My parents, Larry and Marsha Butler, have always been supportive throughout the many endeavors of my life, even when my choices may have seemed less than ideal. I thank you both for your longstanding love and support. My brother

James and my sister Annie have also always been supportive, and have been there to provide a dose of laughter when I most needed it. You are both an inspiration to me, and I feel lucky to have you both in my life. I would like to thank my aunt Edith, whose open-mindedness and sense of wonder has been an influence on me for as long as I can remember, and which helped prompt me to explore the many ideas that have shaped my understanding, eventually culminating in the thoughts presented here in this work. I would like to express an extra special thanks to my wife Julie, who has been with me throughout the ups and downs of this project. Your support has been crucial to this dissertation in a number of ways, and I will always cherish your love, understanding, and companionship. Last but not least, I would like to thank my daughter Lily, who was born not long before my work on this dissertation began. You have been an inspiration to me each and every day, and I will always treasure the memories I have of you being by my side while writing this dissertation.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iv

TABLE OF CONTENTS vi

ABSTRACT ix

CHAPTER 1 – INTRODUCTION 1

CHAPTER 2 – INTROSPECTION AND PERCEPTION 5

2.1 Introduction 5

2.2 Introspection as Perception: An Overview 6

2.3 Poking Out the Inner Eye 17

2.4 What Are the Objects of Mental Perception? 22

2.5 How Do Mental Objects Appear to Us? 30

2.6 Perception and Recursion 32

2.7 What Is Perception Anyway? 33

2.8 Introspection and Attention 43

2.9 The Prominence of Perceptual Metaphor 46

CHAPTER 3 – THE DIVERSITY OF INTROSPECTION 52

CHAPTER 4 – KNOWING OUR OWN CONSCIOUS STATES 55

4.1 Introduction 55

4.2 Knowing What It Is Like 56

4.3 What Kind of Knowledge Is This? 60

4.4 Objections and Clarifications 66

4.5 Putting Skepticism of Self-Knowledge in its Place 73

4.6 Conclusion 80

CHAPTER 5 – INTROSPECTION THROUGH FUNDAMENTAL COGNITIVE PROCESSES 82

5.1 Introduction 82

5.2 Representation, Self-Representation, and Metarepresentation 83

5.3 Conceptualizing Our Own Mental States 96

5.4 Attention and Introspection 104

5.5 Conclusion 111

CHAPTER 6 – UNDERSTANDING OUR OWN BELIEFS AND DESIRES: THE ROLE OF FOLK PSYCHOLOGY IN SELF-UNDERSTANDING 113

6.1 Introduction 113

6.2 The Common-Sense View 114

6.3 The Theory Theory 117

6.4 Empirical Support for the Theory Theory 123

6.5 Mind-Monitoring Mechanisms and the Phenomenology of Propositional Attitudes 130

6.6 Epistemic Implications 137

6.7 Conclusion 141

CHAPTER 7 – THE INTERNAL MONOLOGUE 143

7.1 Introduction 143

7.2 Knowing Our Own Thoughts through Language 145

7.3 Objections and Replies 159

7.4 Kinds of Self-Knowledge Enabled by Inner Speech 171

7.5 From Self-Determined Truth to Self-Deception and Back: The Epistemology of Inner Speech 175

7.6 Conclusion 185

CHAPTER 8 – CONCLUSION: THE NATURE, SCOPE, AND LIMITS OF INTROSPECTION 188

BIBLIOGRAPHY 194

ABSTRACT

In this dissertation I offer a new framework for understanding introspection, characterizing it as a multi-faceted phenomenon that has a broad range of epistemic qualities. I begin by arguing that the standard understanding of introspection as a kind of perception through which we observe our own minds is misguided as a literal account of introspection. With this established, I move on to discuss our diverse abilities to know, and fail to know, our own minds. First, I describe the uniquely first-person experiential knowledge that we have in virtue of being constituted by our own conscious mental states. Second, I illustrate how we know about our own minds through higher-order cognitive abilities, through which we represent, conceptualize, and pay attention to aspects of our own minds in much the same way that we do in understanding the world around us. Finally, I present and discuss several different ways in which language, through the phenomenon of inner speech, contributes to our ability to introspect. Throughout this project, one prominent goal is to map out the varied epistemic dimensions of introspection. Some kinds of introspection provide unique ways of knowing our own minds that are not subject to typical epistemic error, such as the experiential self-knowledge intrinsic to being in a conscious state. However, many other kinds of introspection are subject to error, such as the inherent capacity to misrepresent anything that is mediated through conceptual and representational thought and the capacity to deceive ourselves through the narratives we construct in inner speech. In addition, we can also engage in self-constitutive

mental processes, such that what we know about ourselves can at times be a function of what we ourselves determine. Taking note of these diverse aspects of introspection provides a foundational framework for further empirical inquiry and conceptual analysis, whereby the broad range of introspective activities we engage in, along with the impacts they have on both our ordinary lives and our intellectual investigations into the nature of the mind, can be more accurately understood, evaluated, and utilized.

CHAPTER 1 - INTRODUCTION

What is introspection? Do we literally perceive our own mental states, or are there other processes at work in our ability to know our own minds? What is the epistemic status of introspection? Can it be trusted as a viable or even privileged source of knowledge, or are there problems and limits regarding what we can know about our own minds from our own first-person perspective? In this dissertation, I will address these interrelated issues. In doing so, I offer a new conceptual framework for understanding introspection, suggesting that it is a varied and multi-faceted phenomenon that cannot be limited to a single cognitive mechanism or epistemic characterization. There are a variety of different ways in which we engage in introspection and, correspondingly, a variety of different epistemic dimensions involved. Through presentation and analysis of these various aspects of introspection, I illustrate how we know, and fail to know, our own minds.

I begin my reevaluation of introspection in Chapter 2 by arguing that the standard, somewhat intuitive, understanding of introspection as a single perceptual faculty through which we observe our own minds is misguided. We do not literally perceive ourselves, inside a so-called mind's eye. Instead of being a literal account of how we know our own minds, I suggest that this view of introspection occurs through the pervasive use of perceptual metaphors in our general understanding of knowledge. This is not to say that we should necessarily stop using the concept of introspection, but rather that we need to realize that what appears to be a single

faculty is instead a heterogeneous collection of different ways of knowing our own minds. With this established, I move on to provide a new framework for understanding introspection, as a multi-faceted phenomenon. Throughout the remaining chapters, I describe several different (but sometimes overlapping) ways in which we know our own minds. First, there is the experiential knowledge that we have in virtue of being ourselves, as conscious beings (or at least collections of conscious processes). This kind of self-knowledge, which I characterize as phenomenal, non-propositional knowledge that is unique to one's first-person experience of being conscious, is the focal point of Chapter 4. Second, there are the things that we know about ourselves through our mind's own higher-order cognitive functions, whereby we can represent, conceptualize, and pay attention to aspects of our own minds in much the same way that we understand the world around us. I introduce these processes as foundational aspects of our introspective capacities in Chapter 5, defending a somewhat reductionist account of introspection, in that it can be explained through cognitive faculties that are already in place, irrespective of any particularly introspective functions they may perform. In Chapter 6 I analyze a more specific element of this domain: the use of the belief / desire framework inherent in our folk understanding of psychological phenomena to understand prominent aspects of ourselves. Finally, in Chapter 7, I present and discuss several different ways in which language, through the prominent phenomenon of inner speech, contributes to our ability to know our own minds. Taken together, these various faculties culminate in producing what we ordinarily understand as introspection.

Throughout my discussion of these various aspects of introspection, one prominent goal is to map out the epistemic dimensions of introspection. Some introspective faculties provide unique ways of knowing our own minds that are not subject to typical epistemic error, such as the experiential self-knowledge intrinsic to being in a conscious state and the self-determining truth values of some types of inner speech utterances (such as "I am thinking that I am awake right now"). However, many other kinds of introspection are subject to error, such as the inherent capacity to misrepresent anything that is mediated through conceptual and representational thought and the capacity to deceive ourselves through the narratives we construct with inner speech commentary upon ourselves and our experiences. Moreover, we can also engage in self-constitutive mental processes, such that what we know about ourselves can be at least partially a function of what we ourselves determine. Taken together, these various points illustrate how introspection has a rather broad range of epistemic characteristics.

By drawing attention to this broad epistemic range of introspection, I hope to clarify the needless and misguided debate concerning whether or not introspection is special and privileged or fallible and untrustworthy. Instead, introspection, as a diverse phenomenon, is spread out across all of these characteristics and cannot be categorically rated along a single epistemic dimension. Realization of this fact is long overdue in the various disciplines concerning the human mind, where perspectives on introspection have been polarized by one-dimensional characterizations, from the categorically dismissive attitude that has been prominent in psychology since the

decline and rejection of the early "introspectionists" to those in philosophy who regard introspection as a privileged and authoritative source of knowledge. Introspection cannot be accurately squeezed into either of these characterizations, and must be regarded as a multi-faceted phenomenon with both trustworthy and misguiding epistemic elements. Noting and characterizing this epistemic diversity is progress in our understanding of introspection, but it is certainly not the final word on the matter. Instead, I see my work here as providing a foundational framework for further empirical inquiry and conceptual analysis, whereby the broad range of introspective activities we engage in, along with the impacts they have on both our ordinary lives and our intellectual investigations into the nature of the mind, can be more accurately understood, evaluated, and utilized.

CHAPTER 2 – INTROSPECTION AND PERCEPTION

The Self-Existent made the senses turn outward. Accordingly, man looks towards what is without, and sees not what is within.

- The Katha Upanishad

How pathetically scanty my self-knowledge is compared with, say, my knowledge of my room …. Why? There is no such thing as observation of the inner world, as there is of the outer world.

- Franz Kafka

2.1 Introduction

The idea of introspection typically suggests an ability to perceive within oneself. If we regard introspection in this literal manner, we are led to posit some kind of perceptual faculty in the mind, through which we come to have knowledge of our own mental states. Just as we perceive the external world through specific sensory mechanisms, such as eyes, ears, and so on, we might perceive internal events through some similar sort of mechanism or process that is specially aimed towards the perception of mental states. A number of people, both past and present, have regarded introspection along these lines. Moreover, ordinary talk of a 'mind's eye" or "looking within" oneself likewise suggests some kind of perceptual process regarding introspection. However, I believe that this view of introspection is misguided, at least as a literal account of what goes on in our minds when we come to know our own mental states. We do not literally perceive our own mental states.

Dismissing this common view of introspection will open the concept of introspection to a broader range of understanding and analysis, regarding it as a metaphor for our rather diverse and heterogeneous capacities to obtain knowledge of our own minds.

2.2 Introspection as Perception: An Overview

Before explaining my reasons for dismissing the perceptual account of introspection, I want to first illustrate how common and prominent this viewpoint has been and, to a large extent, continues to be. I will do this by surveying in broad scope a variety of people that all take up this idea in one way or another. My purpose here is not to give a full and definitive account of these viewpoints, but rather to note the breadth and commonality of the introspection-as-perception view.

First of all, a number of historically significant philosophers can be reasonably interpreted along these lines. Although some earlier antecedents might be noted, such as Aristotle, Augustine, and Aquinas (Lyons 1986, pp. 1-2; Wallace 2000, p. 76), I will begin with John Locke. After discussing the perception of sensible objects via ideas in the mind, Locke goes on to say:

> Secondly, The other Fountain, from which Experience furnisheth the Understanding with *Ideas*, is the *Perception of the Operations of our own Minds* within us, as it is employ'd about the *Ideas* it has got; which Operations, when the Soul comes to reflect on, and consider, do furnish the Understanding with another set of *Ideas*, which could not be had from things without: and such are, *Perception, Thinking, Doubting, Believing, Reasoning, Knowing, Willing,* and all the different actings of our own Minds; which we being conscious of, and observing in ourselves, do from these receive into our Understandings, as distinct *Ideas*, as we do from Bodies affecting our Senses. This

> Source of *Ideas*, every Man has wholly in himself: And though it be not Sense, as having nothing to do with external Objects; yet it is very like it, and might properly enough be call'd internal Sense. But as I call the other *Sensation*, so I call this *REFLECTION*, the Ideas it affords being such only, as the Mind gets by reflecting on its own Operations within it self. (Locke 1975, Book II, Chapter I, section 4, p. 105).

So, according to Locke, we perceive the inner workings of our minds through some kind of observational faculty. Although he says it is not "Sense" (which he reserves for referencing our perception of the external world), he thinks it is similar enough to ordinary sense perception to warrant calling it "internal Sense." As we will see, clarifying the manner in which introspection is like sensory perception is a crucial issue, but unfortunately Locke says little else about it. Yet, from the above passage, it is clear that he regards the connection in a somewhat literal manner. He speaks of *observing* the operations of the mind, *perceiving* them as ideas in the same way that we perceive external objects as ideas (under the empiricist theory of ideas, of course). There is no hint that this language is metaphorical, and the only reservation about calling introspection a kind of "sense" is in regard to distinguishing between internal and external objects. So, Locke provides one example of understanding introspection in literally perceptual terms. Similar accounts can be found in Descartes, Hobbes, Hume, and perhaps others (Lyons 1986, p. 2,3), indicating that the idea that we can perceive the workings of our own minds was common, and perhaps even simply assumed, in the 17th century.

This is partially confirmed by looking at Kant, who inherited and transformed many of the ideas of the 17th century, both from the empiricists and the rationalists.

In the *Critique of Pure Reason*, Kant repeatedly speaks of an "inner sense," as a faculty by which we perceive mental phenomena. For instance, in regard to perceiving the self, Kant states that

> The relation of sensibility to an object and what the transcendental ground of this [objective] unity may be, are matters undoubtedly so deeply concealed that *we, who after all know even ourselves only through inner sense and therefore as appearance*, can never be justified in treating sensibility as being a suitable instrument of investigation for discovering anything save always still other appearances… (Kant 1985, A278 / B334, p. 287) italics added

In simpler terms, the idea is that sensory objects can only be known as appearances (rather than as they are "in themselves"), and that this applies not only to external objects, but also to internal, mental objects as well, including our own self. So, just as we come to know of tables, chairs, trees, people, and so on through the appearances they produce in our senses, we likewise come to know of our own mental lives by their appearing to us through some kind of inner sensory modality. In claiming that this appearance / reality distinction is equally applicable to objects of inner and outer sense, Kant clearly regards these as very similar sensory modalities. While there may be some difficulties in translating Kant's own idiosyncratic language over to our now common talk of introspection, the similarities are clear enough to show that Kant, like Locke, thinks that we obtain knowledge of our own minds in some kind of perceptual manner.

Moving along in history, another philosopher (and psychologist) who regards introspection as a kind of perception is William James. In describing his methodology for psychology, James states that:

> *Introspective observation is what we have to rely on first and foremost and always.* The word introspection need hardly be defined – it means, of course, the looking into our own minds and reporting what we there discover. *Every one agrees that we there discover states of consciousness.* (James 1890, vol. 1 chapter VII)

This is the most confident statement we have seen so far that introspection is a kind of perception. James thinks that it is so obvious that introspection is a kind of mental perception that it need not even be explained, and regards this obviousness as so fundamental that it becomes his foundation for psychological research. James is illustrating a very common intuition here, simply taking it for granted that we can perceive our own minds. I will later argue that this view of introspection is not nearly as obvious as it seems, and that what we have here is not a literal understanding of introspection, but rather the persistent application of perceptual metaphor. For the time being, however, the point to notice is how obvious and basic it seems to James that introspection is a kind of perception. To James, and to many others, it is simply given that we observe our own conscious mental states through a perceptual faculty called introspection.

The idea of introspection as a kind of perception was put to practical use by a group of early psychologists that are now referred to (in a somewhat pejorative manner) as the "introspectionists" (Lyons 1986, pp. 3-6.; Danziger 1980). Most notably, this includes the work of Wilhelm Wundt and E. B. Titchener. Like James, these introspectionists took introspection, as the inward perception of mental states, as a foundational method for psychological research. Unlike James, however, they tried to narrow the focus of introspection, as an actual perceptual faculty, so that it

could operate somewhat like a scientific instrument. By training experimental subjects to isolate their introspective observations, the introspectionists sought to uncover the basic constituents of experience, much as chemists documented the basic constituents of physical substances via the periodic table. However, as those familiar with the history of psychology will know, different introspection-based research projects came up with different, and even opposed, results. Such conflict paved the way for other foundational psychologists to dismiss the use of introspection, and the very different behaviorist approach came to prominence. Even though behaviorism has also since gone by the wayside (for the most part), its rejection of "introspectionism" as a psychological approach is something that can still be found in psychological research today, with almost univocal prominence being given to third-person observation and analysis. Nevertheless, the idea that introspection is a kind of perception is far from dead, even among psychologists (Humphrey 1986; Neisser 1993; Silvia and Gendolla 2001). It is worth noting here that this does not necessarily imply any kind of inconsistency. It is possible to hold that introspection is a kind of perception without thinking that such perception should be the basis for psychological research. Perhaps, for instance, it is a particularly inaccurate or biased kind of perception, and therefore cannot be a trustworthy source of information. Such a view of introspection can be found in at least some recent work in psychology (See Silvia and Gendolla 2001, for instance). Moreover, mention of "self-perception" can be found within a variety of contexts in contemporary psychological research. This often occurs with little or no conception or elaboration regarding what this "self-

perception" is supposed to be, in terms of actual, concrete processes. Rather, it seems to be simply assumed (just as we saw above with William James) that people access their own psychological states in some perceptual manner. So, despite the demise of introspectionism in psychology, the view that humans understand themselves through some kind of internal mental perception still persists in at least some areas of contemporary psychology.

I will now turn to some contemporary philosophers who defend the claim that introspection is a kind of perception. I want to remind the reader again that this is only meant to be a cursory survey of the viewpoints I'm addressing. As I go on later to discuss specific aspects of introspection, various features of these accounts will be considered, evaluated, and criticized in more detail. For the time being, though, my purpose is simply to provide a 'big picture' perspective of the introspection-as-perception viewpoint.

First up is David Armstrong. In explaining his account of introspective consciousness, Armstrong utilizes the example of a long-distance truck driver. He states that "After driving for long periods of time, particularly at night, it is possible to "come to" and realize that for some time past one has been driving without being aware of what one has been doing." (1997, p. 723). The idea is that such a person, prior to "coming to", is lacking introspective consciousness. He is perceptually aware of the environment and his actions, at least insofar as such awareness is required for driving, but he is not aware of his awareness. For Armstrong, this lack of awareness is a lack of perceiving one's own mental states. In illustration of this, he states

> What is it that the long-distance truck driver lacks? I think it is an additional form of perception, or, a little more cautiously, it is something that resembles perception. But unlike *sense*-perception, it is not directed toward our current environment and/or our current bodily state. It is perception of the mental. Such "inner" perception is traditionally called introspection, or introspective awareness. (1997, p. 724)

Like Locke, Armstrong distinguishes between ordinary sensory perception and introspection while still regarding them both similar enough to justify calling introspection a kind of "inner" perception. So, although he admits of some caution on this point, it is fairly clear that Armstrong adopts the introspection-as-perception viewpoint. In further support of this, consider a passage where Armstrong affirms the Kantian viewpoint we looked at earlier:

> I believe that Kant suggested the correct way of thinking about introspection when he spoke of our awareness of our own mental states as the operation of 'inner sense'. He took sense-perception as the model for introspection. By sense-perception we become aware of current physical happenings in our environment and our body. By inner sense we become aware of current happenings in our own mind. (1968, p. 95)

So, like Kant and others, Armstrong thinks that we come to know our own mental states by perceiving them, just as we come to know about external objects through our various senses. However, unlike the other philosophers we have looked at so far, Armstrong is a materialist and devotes much of his work to understanding the mind in terms of physical processes. Following this general orientation towards the mental, Armstrong suggests that introspection, as a physical process, occurs through a self-scanning process in the brain: "… it will be a process in which one part of the brain scans another part of the brain. In perception the brain scans the environment. In

awareness of the perception another process in the brain scans that scanning." (1986, p. 94). Through this materialist viewpoint on the introspection-as-perception account, Armstrong sees introspection as a contingent, fallible process. Just as our eyes can go wrong in perceiving the world around us, so can our brain scanner go wrong in perceiving our own mental states. So, from Armstrong's perspective, it seems that the only difference between introspection and ordinary sense perception is in terms of what they are directed towards. One is directed inward, while the others are directed outward, but otherwise they are the same sort of physical process.

Paul Churchland has a similar view of introspection (1984; 1985). He states that

> … self-consciousness, on this view, is just a species of perception: *self-perception*. It is not perception of one's foot with one's eyes, for example, but is rather the perception of one's internal states with what we may call (largely in ignorance) one's faculty of introspection. Self-consciousness is thus no more (and no less) mysterious than perception generally. It is just directed internally rather than externally. (1985, p. 74)

Like Armstrong, and most other contemporary philosophers, Churchland is a materialist, and so it may not seem surprising that he identifies introspection as a perceptual faculty in the brain. However, Churchland is a materialist of a specific sort: an eliminative materialist. He seeks not only to explain, but to replace, the mental in terms of the physical. In other words, rather than attempting to show how our ordinary psychological concepts can be understood as physical processes, Churchland suggests that we do away with our ordinary folk psychology altogether, ultimately replacing it with neuroscience (1981). Considering this, I find it rather odd

that he uncritically follows the common conception of introspection as a kind of perception, especially when he acknowledges, as seen in the quote above, that we are generally ignorant of what exactly this perceptual faculty is. As I will elaborate upon later, I think that this ignorance ought to prompt some skepticism regarding the introspection-as-perception viewpoint. Yet, Churchland does not even seem to consider this possibility. Perhaps, like James, he simply assumes it to be an obvious truth. Whatever the case may be, the fact that an otherwise radical philosopher like Churchland uncritically accepts the introspection-as-perception viewpoint gives further testament to the commonality and prominence of this account.

Another noteworthy philosopher that explicitly endorses the perceptual view of introspection is William Lycan. Following Locke, Kant, and Armstrong, Lycan argues that we have the ability to perceive internal states through some kind of monitoring process. Echoing Armstrong, he states that:

> …to be actively-introspectively aware that P is for one to have an internal scanner in working order that is operating on some state that is itself psychological and delivering information about that state to one's executive control unit. (1987, p. 72)

Lycan goes further than some of the other accounts we have considered and claims that this introspective ability accounts for the nature of consciousness, stating that "consciousness is a perception-like second-order representing of our own psychological states and events." (1996, p. 13). However, for our purposes we can set that further claim aside, and focus on the prior idea that we have such a perception-like capacity.

Lycan and the other investigators we have considered thus far think that we possess a perception-like capacity to monitor our internal states, and many of them simply take this as an obvious feature of human mentality. To me this view seems wrong. We do not literally perceive our own mental states and events. I will argue that this account of introspection is only a metaphor for understanding our diverse introspective capacities, and not a literal account of what actually happens in our minds when we introspect.

Before turning to my objections to the self-perception account, I want to mention a couple of additional examples that will drive home the point that this account is a very common and deeply-embedded feature of how people understand their own mental lives. Thus far, all of the examples I have given come from thinkers in the Western cultural tradition. Yet, the idea that we can perceive our own mental states can also be found in other systems of thought. For example, Buddhism takes up the idea that we have a sixth sense, by which we internally observe the workings of our minds (deCharms 1998, p. 66). This idea is often mentioned in a practical context, in describing meditation practices that involve observing mental events. For instance, in explaining Tantric Yoga practice, it is stated that "The primary concern of these teachings is inward contemplation and introspection to directly perceive, in the atmosphere of meditative settledness, the functioning of the mind." (Powers 1995, p. 244; see also Thera 1962). The fact that the introspection-as-perception viewpoint can be found in this very different context gives further indication of its prominence.

One final example of the introspection-as-perception view is represented by the cover image of a recent popular science magazine:

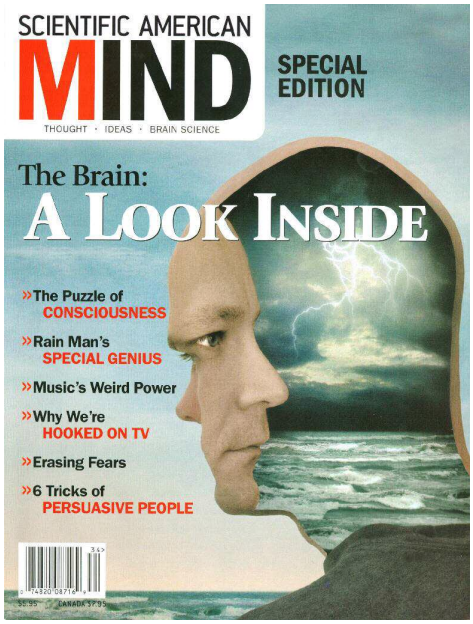

Figure 2.1 – cover image from *Scientific American Mind* (special issue) Vol. 14, No. 1, 2004

We see here an image of a man looking inside his head, with the caption "A Look Inside", suggesting an ability to perceive the contents of one's own mind. While this cannot be taken as a serious account of how we understand our own minds, it does indicate how this idea persists in contemporary popular culture. The idea of "looking within" ourselves through the "mind's eye" is a common, typical, and perhaps even near-universal way of understanding the human capacity for introspection. Yet, as I will now argue, it is mistaken as a literal account of introspection.

2.3 Poking Out the Inner Eye

The first point that I want to make is that there is no identifiable organ through which humans observe the contents of their own minds. Perception is typically associated with a particular organ through which objects and events are perceived: through eyes we perceive visual phenomena, through ears we perceive auditory phenomena, through skin we perceive tactile phenomena, through the tongue we perceive tastes, and through the nose we perceive smells. So, in all of these cases of perception, which encompass the full spectrum of ordinary sensory perception, there is a specific organ that enables the perceptual capacity. However, this appears not to be the case with introspection. Through all the probing and manipulation of human brains that has occurred in the neurosciences, nothing remotely like a sensory organ that takes brain events as input has been identified. So, it seems, there is no empirical basis for regarding introspection as a kind of perception.

In response to this line of criticism, B. Alan Wallace states that:

> This point should come as no surprise, however, in light of the fact that most of the higher order mental functions are little understood by contemporary neurophysiology and there is no cogent neuroscientific understanding of the production of consciousness. … within the brain sciences there is presently a general consensus that, while there is evidence for some sort of localization, there are no precise and different locations or even different sorts of brain tissue corresponding to different thoughts, sensations, or even different types of thoughts or sensations. Thus, the fact that no organs of introspection have been discovered can hardly be counted as grounds for rejecting its existence. (sic) (Wallace 2000, p. 87)

So, although Wallace acknowledges that no organ of introspection has been identified, he thinks that we can still regard introspection as a kind of perceptual faculty, presuming that we just have not yet been able to identify the physical basis for it in the brain. In itself, this is not a good reason for thinking that introspection is a kind of perception. The problem here is that a lack of information on a topic simply cannot provide any positive basis for claims regarding the topic. Just as a lack of information regarding extraterrestrial beings does nothing to support their existence, a lack of understanding regarding the brain does nothing to validate the idea that there is an introspective perceptual faculty. So, Wallace's response should carry little weight.

Now, some will be quick to add that this point equally applies to the claim I am making. The fact that we have not found an introspective perceptual faculty in the brain does not necessarily mean that there is not one. Fair enough. Perhaps the scientific exploration of the brain has simply not developed far enough. Perhaps there is some distributed neural network that operates as a perceptual faculty; we just have not yet found it. Of course, future research always holds the possibility of uncovering some previously unknown or unrecognized object or process. But, for my money, holding out for some future discovery is a risky and groundless enterprise. If we are going to give explanatory weight to introspection as a genuine perceptual faculty, then I think we need at least some positive understanding of what that faculty is and how it works. We simply do not have any such positive understanding of a faculty in the brain that serves the function of perceiving mental states. This purely

negative point alone is not enough to dismiss the perceptual account of introspection, but if it is coupled with the positive grounds for doubt that I will offer, the overall case against this account is strong. If there is no positive evidence of an introspective perceptual device, and there are additional reasons for doubting that introspection occurs in a perception-like manner, then I contend that the introspection-as-perception view ought to be dismissed.

However, a slightly different response to my negative point is to deny altogether the need for a perceptual organ in the brain to account for inner perception. As we saw earlier, Armstrong appeals to the idea of a brain scanner in explaining the nature of introspection. He believes that there is some sort of scanning function performed by the brain upon brain states. Yet, he does not think that this scanning function requires some specific organ to do the job. In support of this, he states:

> It is sometimes argued that introspection cannot be compared to sense-perception because no sense-organ is involved. We say 'I see with my eyes', but there is nothing with which I can say that I discover I am thinking. Now I do not believe that this objection would carry much weight, even if the difference from sense-perception could be made out. But in any case there is one sort of sense-perception where we do not say that we perceive with anything: bodily perception. When I become aware that I am hot, or that my limbs are moving, and I do not gain this knowledge by touch, there is no organ that I can say that I perceive these things *with*. Yet any psychology textbook would take these to be cases of sense-perception. (1968, pp. 95-96)

The problem with this response is that any psychology or human physiology textbook would point to specific physiological processes that are involved in bodily perception. There are actually a number of different physiological processes whereby information about one's own body is processed, such as heat and pain receptors, balance

detection, detection of muscle position and movement, and so on. In all such cases (many of which fall under the general category of proprioception), there are specific sorts of nervous tissues that can be functionally and physiologically identified with the perceptual processes. Now, these specific physiological processes may or may not be counted as perceptual organs, depending on how the concept of a perceptual organ is characterized, but they are clearly physical processes that can be precisely identified with the modes of perception at hand. We have nothing like this when it comes to introspection, as a special and unique kind of perception. There are no identifiable nervous tissues that serve the specific purpose of detecting mental states (whatever that may mean… more on this in the next section). So, if there is no specific physiological process that we can associate with introspection, then how can we legitimately suppose that there is some actual perceptual process involved? It is simply an ad hoc hypothesis, with no empirical founding. To be clear, I am not denying that introspection, whatever it may be, has a physiological basis. Instead, my point here is only that there is no physiological process that can be uniquely identified with introspection, as a kind of perceptual process.

To elaborate on this, let us consider the nature of functional processes. To characterize introspection specifically as a perceptual process is to identify it with a specific function: the function of observing mental states. As many philosophers have noted, functions are "multiply realizable;" a particular function can be performed through diverse sorts of processes. However, in order for a function to actually occur, there must be some embodied process through which it occurs,

irrespective of the point that it could have occurred in some different manner. In other words, if a particular function is performed there is, by necessity, some particular process through which it is performed. Yet, this seems to be what Armstrong is denying. He says that introspection occurs through some sort of brain scanning process, but then denies that there is any actual brain scanner in the brain. This appears to be a fundamental inconsistency in Armstrong's account. If introspection is going to be identified with some perceptual process in the brain, then it seems to me that some such perceptual process and its associated physiological mechanism needs to be identified. Since no such brain-observing process can be identified, I find the perceptual account to be lacking.

Admittedly, just noting the lack of an identifiable organ or physiological process is itself not a strong case against the perceptual account of introspection. It provides some prima facie ground for doubt, but more is needed to thoroughly dismiss the account. So, I will now turn to some further considerations that illustrate the problematic nature of the introspection-as-perception account. Taken all together, these considerations will show that the idea that introspection is literally a kind of perception ought to be left behind.

2.4 What are the Objects of Mental Perception?

One key aspect of perception is having objects of perception. Our eyes see colors and shapes; ears hear sounds; noses smell smells; proprioceptive nerves feel pains, pressures, and other bodily processes (if you count proprioception as a kind of perception… see O'Shaughnessy 1995). If there is a literal perceptual faculty in the mind, such as a "mind's eye", then what sorts of objects does it perceive? Presumably, the answer must be either brain states or mental states (I am not assuming any sort of dualism here… I just want to cover all the potential bases). So let us consider each option in turn.

If the supposed internal perceptual faculty perceives brain states, then these brain states must be "scrambled" in some way, as they do not appear to us in introspection as brain states (Lyons 1986, pp. 57ff). Brain states are incredibly complicated electro-chemical events among innumerable neural networks. However, this is definitely not what we perceive, if we perceive anything at all, through introspection. The thought that 2 +2=4, for instance, does not appear to us as a particular neural firing, or even as any particular physical process at all. So, if we perceive a brain state when we are aware of having such a thought, then that brain state must be filtered through some sort of process that transforms it into something that appears quite different. Otherwise, philosophers would not have had to fuss over the mind / body problem all these years, and could have readily identified mental states as physiological processes! So, if we have the capacity to perceive brain states,

it must be through some mechanism that alters their appearance so radically that they do not even seem like particular physical processes. But, it seems highly improbable that there would be some such scrambling mechanism in the brain, serving the specific function of altering perceptions beyond any recognition as a physical event. Outside the construction of skeptical scenarios, such as Descartes' evil demon hypothesis, there is no reason to suppose that nature, or God, or whatever, would impose such a device upon human cognition. So, this brain scrambling idea is nothing more than an ad hoc hypothesis, having no merit or evidence beyond filling in the gaps of an implausible theory.

To further support my point here, let us consider the other side of the coin. Not only is it implausible to think that introspection is the perception of brain states; it is equally implausible to hold that perceiving brain states counts as introspection. Suppose, for instance, that you cut open my head or place me in a MRI in such a way that I can observe my own neurophysiology. It seems absurd to say that I would be introspecting in this case. Even if I had a complete understanding of the brain, being able to accurately correlate specific neural events with specific subjective experiences or cognitive states, I would not be experiencing or understanding my mentality in the same way that I do through introspection. No matter what kind of perception we imagine here, whether it be looking directly at a brain, observing an MRI or other mechanical monitoring process, or even listening to auditory representations of neural firings, it does not seem remotely like introspection in any way. If no other way of perceiving brain states seems at all similar to introspection, we are left with little

reason for thinking that introspection is the perception of brain states. While this does not conclusively show that no perceptual process at all can count as introspection, it leaves the idea rather doubtful.

One reply to this line of criticism might be to point out the possibility that we do not perceive brain states, but rather only representations of brain states. When we perceive ordinary objects, such as tables, trees, and so on, we do not perceive their true physical states (i.e. their atomic structure), but rather various representational features of those states, such as colors and shapes. Analogously, perhaps we perceive our brain states through representational features of those states. But what would these representations be? Clearly they would be mental states, such as pains, thoughts, beliefs, desires, and so on. So, while we can rule out the direct perception of brain states for the reasons given above, perhaps there is room left to say that we perceive such mental states.

As anyone familiar with the philosophy of mind will know, there are numerous ways of thinking about the nature of mental states and their relations to brain states. Perhaps mental states are simply identical to brain states. Alternatively, we might categorize mental states as being somehow separable from brain states, for either epistemic or metaphysical reasons. This separation could be understood along dualist lines, but need not be. Other perspectives can also support separate understandings of mental states and brain states, such as anomalous monism, nonreductive physicalism, emergentism, dual aspect theory and so on (for an introductory survey of the options here, see Kim 1998). In any case, let us assume, for

the sake of argument if nothing else, that mental states do exist in some manner or another, and consider whether or not they could be objects of inner perception.

While there are numerous kinds of mental states, let's narrow our scope a bit and focus on two prototypical examples of mental states: emotions and beliefs. If we can perceive mental states, then surely these two staples of mental life would be included among the kinds of things can perceive. So, can these be counted as objects of perception, in the same way that colored objects that move and make sounds are perceptual objects? It doesn't seem to me that they can. First of all, beliefs and emotions do not have the objectively identifiable unity possessed by externally perceived objects (for those familiar with the concept, what I have in mind here roughly corresponds to perceptual constancy). For instance, it is notoriously difficult to identify *what* exactly a belief *is*, and to identify precisely *where it is*. Arguably, the same could be said of emotions as well. These mental states seem to lack the easily identifiable perceptual properties possessed by typical objects, such as trees, animals, cars, rocks, and so on. Nor do they seem to persist through time as single, isolatable objects, as do pencils, seashells, and other ordinary objects. It is difficult to say when and where a belief begins and ends, and to say how to separate beliefs from one another. Consider, for instance, your belief that you are reading the word "now" now. Is that belief the same belief that you are having *now*, or is there *now* a new belief? At this point, what has happened to that belief? You are no longer reading the word in question, so presumably you no longer have the belief. But if you go back and read that sentence again, there again is the belief. Or is that a new belief? And if that

belief is a perceptual object, what is it that you are perceiving, exactly? I hope that my point is clear. Beliefs, if they are objects at all, are amorphous, changeable, and protean. They do not have the sorts of properties that make an object perceptually identifiable.

Many have argued that beliefs must be defined purely relationally, in terms of the functions, or conceptual roles, they perform through interacting with other mental states, rather than as objects that can be defined and understood in terms of inherent properties. I tend to agree with this perspective, though it would take us off target for me to explain why here. If this is the right way to think of beliefs (or, if it is even a plausible option), then beliefs could not be perceptual objects, by their very nature. They are informational relations, and not objects in themselves. Consider again some belief that you have about what you are reading. From this perspective, that belief *is the belief that it is* in virtue of how it functions in your cognitive economy- e.g. the role it plays in your understanding of what you are reading, its role in your ability to relate what you are reading to other information you have, and so on. Beliefs, then, are not objects that can be perceived. They are postulated, inferred, assumed, or hypothesized, based on the processes in which they function. Beliefs are more like gravitational forces than falling objects.

To take these worries about perceptual objects a bit further, let me outline a further possible perspective here. Perhaps mental states, such as beliefs and emotions, should not be thought of as objects, but rather as subjective events that engender the possibility of having perceptual objects. Following the

phenomenological tradition of Husserl, many have suggested that we think of the self as a subject rather than an object (for a recent illustration of this idea, see Zahavi 2002). From this perspective, the focus is on the self as subject, as the possibility for experience, rather than an object of experience itself. Along these lines, perhaps mental states ought to be thought of as events that we are constituted by, as cognitive and experiential subjects, rather than as things that we experience as objects. In other words, perhaps we do not know about mental states as objects in the way that we do the objects of the so-called external world. Instead, we know about them by *being* them.

John Searle offers a similar perspective, arguing against the idea of introspection as a kind of inner observation on the grounds that conscious mental states are inherently subjective. He states that:

> … where conscious subjectivity is concerned, there is no distinction between the observation and the thing observed, between the perception and the object perceived. The model of vision works on the presupposition that there is a distinction between the thing seen and the seeing of it. But for "introspection" there is simply no way to make this separation. Any introspection I have of my own conscious state is itself that conscious state. (1992, p. 97)

I think that there is something to Searle's point here, but it goes a bit too far. There is a significant incongruity between ordinary objects and mental states. The former can be objects of perception while the latter cannot. We can perceptually observe trees, birds, chairs, and so on, but when it comes to beliefs, desires, personality traits, and such, we are at pains to say in what way these things could even count as perceptual objects. Contra Searle, however, this is not to say that a particular mental state cannot

be understood separately from the subjective experience of the person who is having it. As I will explain in more detail later, mental states can become objects by becoming conceptualized in a theory-like manner. In addition to having, or being constituted by, mental states, we also learn about them by conceptualizing and thinking about them, just as we do with other, extramental things. In effect, this turns mental states into objects of a sort, just as analyzing physical movements leads to positing forces as objects of inquiry. But, these are not objects that we perceive. Rather, they are theoretical posits, allowing us to make sense of certain phenomena. For instance, suppose you pause a moment to reflect upon what you are reading. In doing so, you will quite likely, irrespective of my suggestions, think in terms of beliefs, such as "I believe that this argument is mistaken" or "I can't believe that this guy believes that." We use the concept of belief quite automatically, as a way of making sense of thought, both regarding ourselves and others. In doing so, we utilize beliefs as objects, not in the sense of things that we perceive, but rather in the sense of things that we posit to make sense of the world. So, while I reject the idea that mental states are perceived as objects through some introspective faculty, mental states can be, and are, treated as objects through our various conceptual capacities.

In this section I have argued that there are no objects of perception involved in introspection – or, at least, that we have no good reason to think that there are. In summation, I want to point to a similar concern posed by Güven Güzeldere, in criticizing higher-order perception accounts of consciousness. He states:

> Locke, Armstrong, Churchland, and Lycan all talk about consciousness as the awareness, or perception, or monitoring, or

> scanning of one's mental states. There is an ambiguity inherent in all these statements concerning the nature of the proper object of such internal, perception-like awareness. What exactly is being perceived in the "perception of what passes in one's own mind": the content of the mental state that happens to be "passing through" one's mind at the time or the mental state itself? Or the content of another thought to the effect that one is having such a mental state? There is surprisingly little attention paid to spelling out the answer to this question in any detail. (1997, p. 792)

Although Güzeldere is concerned here with the explanation of consciousness, we can set that aspect aside and focus on the way the accounts he is discussing treat introspection as a kind of mental perception. The point to take home from this quote is that such accounts lack any clear explanation of what exactly is involved in such perception. As Güzeldere points out, none of these accounts cash out how this perception takes place or what it involves. Rather, it is simply assumed that mental states can be objects of perception. Presumably, the authors of these accounts think that it will be obvious to their readers what is perceived and how the perception takes place. I hope to have shown in this section that this presumption cannot be made. There is a reason that perceptual accounts of introspection have little to say about *what is actually perceived*, and it is not because *what is actually perceived* is obvious. Rather, it is because there is no reasonable way to construe mental states as perceptual objects. We simply do not perceive beliefs, pains, desires, fears, or personality traits.

2.5 How do mental objects appear to us?

If, contrary to what I've just argued, there are mental objects that can be perceived, what are their phenomenal qualities, or qualia? One thing that perceptual objects have in common is that they have distinctive phenomenal traits. They all "appear" in their own idiosyncratic way, each having a unique manner in which there is "something that it is like" to perceive it. Philosophers have used the term "qualia" to refer to these phenomenal traits. Think again of the way visual objects appear, in terms of color, shape, and movement, or the way sounds produce their particular kinds of experiences. Do mental states have comparable phenomenal traits, or ways of seeming? Sure, emotions have particular sorts of *feels*, but not in the same vivid, unique way that phenomenal objects do. Emotions, insofar as they pertain to perception, are modes of perceiving, rather than perceivable objects in themselves. Fear, for example, is not something we perceive in ourselves, but rather something that we *feel* in association with perceiving or thinking about something else. If you don't quite follow me here, perhaps it will help to consider beliefs instead. Arguably, beliefs have no way of feeling at all. Perhaps there is some sense of confidence associated with having a belief, but a belief is not itself an object that we perceive through some phenomenal way of seeming. Even David Chalmers, an adamant champion of conscious phenomenal states, is skeptical of the idea that beliefs have some distinctive sort of qualia or phenomenal appearance through which they may be said to be beliefs. Here is what he says about it:

> Certainly, there is often conscious experience in the vicinity of belief: there is something it is like when one has an occurent (i.e. conscious) belief, and most nonoccurent beliefs can at least bring about a conscious belief. The crucial questions, though, are whether this conscious quality is what *makes* the state a belief, and whether it is what gives it the content it has. This may be more plausible for some beliefs than for others: for example, one might argue that a conscious quality is required to truly have beliefs about one's *experiences*, and perhaps also certain sorts of experiences are required to have certain sorts of perceptual beliefs about the external world (perhaps one needs red experiences to believe that an object is red?). In other cases, this seems more problematic. For example, when I think that Don Bradman is the greatest cricketer of all time, it seems plausible to say that I would have had the same belief even if I had had a very different conscious experience associated with it. The phenomenology of the belief is relatively faint, and it is hard to see how it could be this phenomenal quality that makes the belief a belief about Bradman. What seems more central to the belief's content is the connection between the belief and Bradman, and the role it plays in my cognitive system. (1996, p. 20)

So, while there might be qualitative, phenomenal experiences associated with beliefs, beliefs are not themselves constituted by some phenomenal way of seeming. Although Chalmers is not concerned with the nature of introspection in making this point (his concern is to distinguish between psychological and phenomenological processes), it bears an important implication for our topic: we do not identify beliefs by the way they phenomenally appear to us, but rather by the roles they play in our cognition.

Later, I will elaborate on this point, in discussing how we come to have knowledge of our own beliefs. For now, in regard to the claim that we do not perceive beliefs through perceptual qualia, all I can really do is appeal to my own experience, and count on the reader having the same sorts of experiences (or, to be more accurate, the lack thereof) that I do. Perhaps I've had my inner eye plucked out,

and I'm like a blind man denying the vivid nature of colors, but it sure seems to me that mental states do not have the same sorts of phenomenal qualities that characterize perceptual experience. They do not present themselves in the same vivid, qualitatively identifiable way that perceptual objects do. There are only our ordinary perceptual experiences of external objects and our conceptual reflections upon them. There are no perceptions of these mental processes as phenomenal objects in themselves. However convincing or unconvincing this may be to others, I add it to my list of reasons for thinking that introspection cannot be literally understood as a perceptual faculty.

2.6 Perception and Recursion

Yet another reason for rejecting the perceptual account of introspection pertains to the fact that introspection is recursive, in that we can introspect our introspections. For instance, I can think about my thought that as a child I once saw a rabid dog running through my neighborhood. I can wonder about whether this thought is veridical or not. Did it actually happen or is it a falsely reconstructed memory? I can further think about this wondering, reflecting upon why I am wondering whether my memory is veridical or not. Perhaps this can only plausibly go a few steps, due to the limits of our cognitive abilities, but it is sufficient to show that introspection is recursive. We can introspect our introspective states. We can think about our thoughts, reflecting upon our mental states, including mental states

about mental states. No perceptual faculty can do this. We cannot hear or otherwise perceive the *hearing of the sounds* that we hear; we only hear the sounds. Our eyes cannot see their own visual perceptions; there are just the perceptions that they, in conjunction with the brain, construct. Contrary to a popular phrase, we cannot feel our pain. There is just the pain itself, as a feeling. Of course, we can think about and reflect upon the pain, or any other perceptual state, but that is not a case of recursive perception. Rather, it is applying some other, more abstract, cognitive process to the pain. So, to summarize, perceptions cannot themselves be perceived, but introspective states can be introspected. Therefore, introspection is unlike perception in this fundamental way.

2.7 What is perception anyway?

One concern that may have occurred to you by now is that I have not offered an account of perception. If I am arguing that introspection is not a kind of perception, then shouldn't I tell you what perception is? I will try to meet this concern here by explaining my understanding of perception. The nature of perception is a huge topic in itself, so it would be impossible to provide a full analysis of it here. Nevertheless, an overview of the topic will help clarify important aspects of my claim that introspection is not a kind of perception. So, I will describe a broadly evolutionary account of perception, incorporating aspects of both the common layman's understanding of perception and cognitive psychology. I will then

show how this account provides additional reasons for doubting the idea that introspection is a kind of perception.

Among the standard cases of perception, there are a number of traits that they all share. I have already noted a few of these: Each kind of perception is associated with a specific organ that performs perceptive tasks; through the various kinds of perception, unique perceptual objects are perceived, with unique, idiosyncratic phenomenal qualities; modes of perception cannot perceive themselves. As we have seen, introspection does not share these properties. I think that the absence of these traits provides good grounds for doubting the perceptual account of introspection. Strictly speaking, however, it does not entirely rule it out. Perhaps introspection is a kind of perception that just doesn't happen to share these traits with other kinds of perception. Perhaps there are other features of perception that introspection has, which justify the idea that introspection is a kind of perception, or at least perception-like in some important respect. To decide whether or not this is the case, we need to understand the essential nature of perception. There is no simple, authoritative account to draw upon here, but I think the following considerations will at least head us in the right direction.

Perception is usually closely tied to sensation. Kinds of perception are typified by the kinds of sensations that an organism has. Humans, for instance, perceive sounds through their ears, colors, shapes, and movements through their eyes, smells through their nose, tastes through their mouth, and textures through their skin. It is this commonsensical way of thinking about perception that I have mainly had in

mind so far. Perception is understood here in terms of a specific organ, or a specific kind of physical process, that takes in information from the environment. As I have already explained, there is no such identifiable organ or physical process that we can specifically relate to introspection. So, introspection cannot be counted as a kind of perception.

On other accounts perception is more closely associated, not with modes of sensation, but with what the mind does with sensory information. From this more cognitive conception of perception, perception is regarded as a form of active information processing (for an overview of this idea, see Pomerantz 2003). Given numerous psychological discoveries regarding the active nature of perception, it is clear that it involves more than the simple sensation of external stimuli. Consider, for instance, the case of visual depth perception. There is a great deal of active information processing involved in the conversion of two-dimensional, and constantly changing, patterns of light impinging upon the retina into the stable three-dimensional panorama that we experience visually (Gregory 1998; Farah 2000). So, we know that perception involves not just the impinging of various stimuli on a receptor, but also a lot of active processing that converts the sensory information into salient data. However, it is important to remember that perception is still very domain-specific. For instance, there are specific perceptual processes dedicated to the processing of visual stimuli that can be accurately correlated with specific areas in the brain (Gregory 1998; Farah 2000; Gaulin and McBurney 2001, pp. 106-109; Palmer and Palmer 2002, pp. 60-65; Jacob and Jeannerod 2003). In other words, there are

identifiable modules dedicated to the various perceptual processes involved in vision. Related to this is the fact that the stereotypical cases of perception involve specific modes of sensation through specific organs. Although our understanding of perception cannot be confined to just the functioning of the sensory organs that are involved, the sensory organs clearly play a crucial role and cannot be ignored.

One way of putting this all together is to understand perception from the perspective of evolutionary psychology. If we think of the human organism, in all of its physical and cognitive aspects, as a product of natural selection, we can try to understand the various specific processes of the organism in terms of evolutionary adaptation. More specifically, perception can be understood in terms of the functional benefits it confers upon organisms in their pursuit of survival and reproduction (Gaulin and McBurney 2001, pp. 91-109; Palmer and Palmer 2002, pp. 60-65). Consider again the eye and its associated cognitive processes, which are now commonly regarded as evolutionary adaptations. From an evolutionary standpoint we can understand vision as a unified functional process, in terms of what it does for organisms that have such processes.[1] Broadly speaking, it allows organisms to detect certain salient features of objects in the environment, such as the color and shape of a specific piece of fruit, the movements of a potential predator, or the face of a

[1] It is important to note that I do not mean "unified functional process" in the sense that there is only one identifiable process involved. It is now clear that there are separate modular processes involved in specific kinds of visual information processing. There are, for instance, separate dedicated modules for detecting color, shape, motion, faces, and perhaps other aspects of vision (Gregory 1998; Farah 2000; Gaulin and McBurney 2001, pp. 106-109; Palmer and Palmer 2002, pp. 60-65; Jacob and Jeannerod 2003). My point here is that these individual processes, while they each constitute a functional unity in themselves, can also be said to more broadly constitute a functional unity in terms of general visual perception. They all work together to integrate visual environmental information into the various purposes of the organisms that have them.

particular conspecific. In performing such functions, the eye works with specific cognitive processes as a functional unit that enables an organism to obtain information about its environment. This functional unit can plausibly be regarded as an evolutionary adaptation. It is a complex collection of tightly knit processes that all work together to promote the survival and reproductive success of the organisms that possess it. In the case of perception, broadly speaking, the adaptive advantage arises from the ability to process salient information about the present state of the world. Setting the details aside, various types of perception may be characterized in terms of how they function to provide information to an organism about the state of the world it lives in. Of course, much more could be said to fully illustrate and defend this view of perception (as well as the more general presumptions regarding adaptations and functions), but I think that this brief overview presents a plausible and useful way of thinking about it.

This leads me to yet another reason for rejecting the idea that introspection is a kind of perception: There appears to be no identifiable functional / adaptive process that serves the purpose of perceiving one's own mental states. This is related to my initial objection regarding the absence of an introspection organ, but shifts the focus away from the organ towards the function it supposedly performs. In order to get a handle on this objection, let us first look at an account that suggests the opposite. Nicholas Humphrey offers the following hypothesis:

> Now imagine that at some time in history a new kind of sense organ evolves, the inner eye whose field of view is not the outside world but the brain itself. Like other sense organs, the inner eye provides a picture of its information field that is partial and selective; but equally

> like other sense organs it has been designed by evolution so that its picture is a useful one, a 'user-friendly' description which tells the subject just so much as he requires to know in a form that he is predisposed to understand – allowing him by a kind of magical translation to see his own brain-states as conscious states of mind. … Suppose that is indeed what *consciousness* amounts to, and that we human beings are the only animals in nature to have evolved this kind of inner eye. What would it mean for our ability to do psychology? To begin with, it would mean that each individual human being would have almost literally a headstart in reading his own mind. No more fussing about like a behaviourist psychologist with 'intelligent guesses' about what lies behind our own behaviour. We would know immediately where the deeper explanation lies - in our own brain-states, which our inner eye reveals. But in practice it would mean very much more: for the explanation we have of our own behaviour could then form the basis for explaining *other people's*, too. We could, in effect, imagine what it's like to be them, because we know what it's like to be ourselves. (1986, pp. 70-71)

Lest you think that Humphrey is just drawing a metaphorical picture here, he goes on to make this inner eye hypothesis the basis for understanding social intelligence. In essence, Humphrey argues that the "inner eye" is an evolutionary adaptation, hypothesizing that it would confer a significant fitness benefit on those who have one. In contrast, I suggest a different account. I think it is more plausible to hypothesize that our introspective capacities are not adaptively beneficial in themselves. Rather, they are by-products (spandrels, or perhaps exaptations) of other adaptive processes that make them possible.

To illustrate this, let us compare the potential benefits of knowing ourselves versus the benefits of knowing others. In terms of evolutionary fitness, it isn't particularly beneficial for me to have accurate representations of my own mental states. Knowing that I'm a cooperative person, for instance, wouldn't add anything beneficial to my interpersonal interactions. Any benefit would already be conferred

by my actual cooperativeness, regardless of whether or not I accurately represent that feature of myself to myself.

Similar reasoning could apply to other types of mental states, such as beliefs, desires, pains, and so on. Being introspectively aware that I believe there is some edible fruit in front of me, for example, would not seem to add any fitness benefit to me, as the belief itself would already confer any benefit that it may have. Considering this, it seems that there is no real evolutionary pressure for developing some inner eye through natural selection. Being able to understand the mental states of others, however, is clearly beneficial. Correctly believing that another person is cooperative, for instance, would quite obviously confer significant benefits to me. It would enable me to engage in mutually profitable interactions with that other person. Similarly, the ability to accurately represent the beliefs and desires of other people would be very fitness enhancing in that it would enable me to anticipate various significant behaviors (mating prospects, potentially negative and positive social interactions, etc.). So, if there is an adaptation here, it is in regard to our understanding of others, and not ourselves.

This idea agrees with the thinking of a number of psychologists who have postulated the presence of a "theory of mind" in humans and some non-human primates. The concept of a "theory of mind" more or less corresponds to what philosophers call "folk psychology." The "theory" consists of a cognitive module adapted to facilitate social understanding and interaction (Premack and Woodruff 1978; Byrne and Whiten 1988; Carruthers and Smith 1996; Ristau 1998). Roughly,

the idea is that selective pressures for the ability to anticipate the actions of others in one's social environment have led to a specific cognitive capacity to attribute beliefs and desires to others. Although this idea is in need of further development and empirical support (for a critique, see Heyes 1998), it seems fairly plausible to me. At the very least, it seems safe to say that evolutionary pressures helped to form our abilities to represent external events, including the mental states of others. But the idea that there is an adaptive advantage that might lead to the selection for an ability to "see" our own internal states seems more doubtful. So, introspection probably draws from the adaptive ability to understand others, rather than being an adaptive ability in its own right.

This idea is further corroborated by at least some evidence in developmental psychology. For instance, it appears that children come to have an understanding of their own false beliefs at the same time that they understand false belief states in others - at around the age of 3 ½ years. This suggests that understanding false beliefs in both oneself and in others has the same cognitive origins (Gopnik 1993; Gopnik and Meltzoff 1997). If Humphrey and others who postulate an inner eye are correct, it would be rather surprising that this is the case. Instead, arguably, we ought to see the development of understanding one's own beliefs before the development of understanding the beliefs of others, because it is presumably that sort of inferential ability that would have prompted the evolutionary selection of an internal observation mechanism. So, on the "inner eye" model, especially as Humphrey describes it, it seems that children would "see" the nature of beliefs within themselves and then later

use that understanding to infer the existence of similar states in others. Since this is not so, at least in the case of false beliefs, the evidence weighs negatively against the inner eye hypothesis.

One other relevant consideration here is that introspection, or at least certain prevalent forms of it, is relatively rare in comparison with outward-directed perception and thought. Arguably, most human attention is directed towards the external world: navigation, object identification and manipulation, social interaction, and so on. People just don't spend a lot of time focusing on their own mental states. Certainly, it does happen (or there would be no phenomenon to explain), but it isn't nearly as prominent as the vast diversity of external concerns. There is, in fact, some empirical support for this. Mihaly Csikszentmihalyi has done a significant amount of research regarding how people spend their time, and one thing he notes about this is that people generally don't like to be alone (1997, pp. 89ff). He states that

> …people in general report much lower moods when alone than when they are with others. They feel less happy, less cheerful, less strong, and more bored, more passive, more lonely. The only dimension of experience that tends to be higher alone is concentration. When first hearing these patterns, many thoughtful persons are incredulous: "This cannot be true," they say, "*I* love to be alone and seek out solitude when I can." In fact it is possible to learn to like solitude, but it does not come easily. If one is an artist, scientist, or writer; or if one has a hobby, or a rich inner life, then being alone is not only enjoyable but necessary. Relatively few individuals, however, master the mental tools that will make this possible. (Ibid., p. 90)

Evolutionarily, this makes sense. Survival and reproduction largely depend upon how an organism relates to the world rather than itself. So, it seems that we ought to be naturally focused on external things. More anecdotally, consider the fact that

inward-focused religious traditions and practices, such as Buddhist meditation, have to impose strong disciplinary measures upon attention in order to cultivate inner-directed practices. Drawing our attention to our own mental states takes a lot of effort. Taking this into consideration, it seems that introspection is not our forte. It is not something that we do easily or automatically, as is the case with all of our other forms of perception, or adaptive processes more generally (in general, adaptations function effortlessly, without need for outside discipline or control). So, it is implausible to think that we have some special part within us that is dedicated to the task.

In summary, there is no real reason to suppose that there is a perceptual faculty specially adapted to perceive one's own mental states. Instead, it is more plausible to think that our ability to understand ourselves is a byproduct of our adaptive, externally directed, abilities to understand our environment, including other people. Admittedly, these evolutionary considerations are somewhat speculative. In constructing adaptive explanations, especially rather abstracted ones such as these, there is the danger of constructing "just-so" stories that do not reflect actual developmental history, but rather our imaginative capacities. Still, I think that these considerations give added plausibility to my account, in so far as they show how it could plausibly fit with our evolutionary history, and how the alternative view has some problems fitting with a plausible evolutionary account. Coupled with the previous arguments I've given, the case against the perceptual account of introspection is quite strong.

2.8 Introspection and Attention

Before wrapping up my thoughts regarding the perceptual account of introspection, I want to consider one more option for such an account. Although I think that this option ultimately fails to support a perceptual understanding of introspection, it will start heading us in the right direction, towards some possibilities that will be developed later on in this work. According to some proponents of the introspection as perception account, inner perception occurs through some kind of attention device being directed towards our own mental phenomena. William Lycan, in particular, states "As I would put it, consciousness is the functioning of internal *attention mechanisms* directed at lower-order psychological states and events." (1996, p. 14) Insofar as this entails positing some special observational device in the brain, there is a significant problem that such an account confronts. The problem is a version of the homunculus worries that regularly creep up in cognitive science and the philosophy of mind, in that it postulates some kind of inner agent within the mind (in this case, an inner observer of sorts that functions by attending to the diverse functions of the mind). The room for such homunculi is limited, both in terms of actual, concrete processes in the mind and in terms of explanatory devices in the sciences of the mind. I tend to lean towards the idea that there is really no room for them at all; Agents are things to be explained themselves, and not things that are explanatory. So, positing some attentional agent in the brain in order to explain how

we attend to our own minds really doesn't get us anywhere. It is tantamount to replacing one problem with another.

Lycan considers and responds to this concern. Following Dennett, Lycan describes the problem as "Cartesian materialism," in that some kind of inner theater is postulated where it "all comes together"; where some audience-like agent sits back and observes the events in the mind (1996, pp. 30ff; Dennett 1991). Lycan accepts this as a legitimate worry, and then goes on to say:

> But it should be clear that the inner-sense view is not per se committed to Cartesian materialism. For even if an internal scanner resembles an internal audience in some ways, the "audience" need not be seated in a Cartesian theater: There need be no *single*, executive scanner, and no one scanner or monitor need view the entire array of first-order mental states accessible to consciousness. Accordingly, there need be neither a "turnstile of consciousness" nor one central inner stage on which the contents of consciousness are displayed in fixed temporal order. An internal monitor is an attention mechanism, which presumably can be directed upon representational subsystems and stages of same. No doubt internal monitors work selectively and piecemeal, and their operations depend on control windows and other elements of conative context. On these points, the inner-sense theory has already parted with Cartesian materialism. (1996, p. 32)

This is indeed a way of avoiding the homunculus problem. However, if this is how we are to regard introspection, then it seems to be a move away from the perceptual account. First of all, there are reasons for separating attention from perception (though they do clearly work together in a variety of contexts). Most notably, there is some neurological evidence that perception and attention (or, if you prefer, primary sensory processing and secondary attentional processing) occur in identifiably different areas of the brain (Fuster 2002; Picton et. al. 2002). Very roughly, the posterior regions of the primate neocortex are devoted to processing

sensory / perceptual data, while the frontal regions seem to involve the more supervisory, monitor-like cognitive domains associated with attention (Fuster 2002, esp. pp. 97, 100, and 102). So, this suggests that an attentional account of introspection will cash out to be different from the perceptual account.

Secondly, fragmenting the mechanisms involved in introspection, as Lycan suggests, is a step away from the common-sense nature of the perceptual account. One of the main intuitive appeals (perhaps even the main appeal) of postulating a perceptual or perception-like faculty in the mind is that it fits with our intuitive sense of unity in our conscious experience. If we do in fact observe our mental lives, it is from the perspective of the single, conscious "me" that we are all intimately familiar with. However, if we think of introspection not as a single mode of perception, but rather as a collection of smaller, less agent-like, attention mechanisms, then it ceases to fit with this commonsensical way of thinking about ourselves. To be clear, I am not suggesting here that we think of ourselves and our introspective abilities as a single, unified thing. (To the contrary!) The point is that our intuitive unity of consciousness is *part of* the idea that we perceive our mental lives through some kind of "inner sense". If we divorce the perceptual account of introspection from this intuitive idea, then it seems to me that we are moving away from the idea that introspection constitutes a unique kind of perceptual access to our own minds. Of course, I see this as a step in the right direction. As will be discussed in the next chapter, introspection is a diverse, heterogeneous collection of processes, and not a single process or even a single type of process. So, despite the fact that Lycan

persists in calling his account a version of the Lockean inner-sense account, I think that his move towards fragmentation is a good idea. Attention does indeed need to be considered as an aspect of introspection, but not in terms of a single attentional device, and especially not in terms of a unique kind of inner sense or perception. As I will next explain, applying the idea of perception here is merely metaphorical, and is not the development of a literal account of what happens in introspection.

2.9 The Prominence of Perceptual Metaphor

Considering all of the arguments above, I think we can safely dismiss the idea that introspection is a kind of perception. But, this leaves us with some questions. If introspection is not a kind of perception, then what is it, and why do so many people continue to talk about it in this way? I will begin looking at the first question in the upcoming chapter, defending the view that introspection is better understood as a heterogeneous collection of different kinds of processes. Regarding the second question, I find it extremely useful to take a look at the work of George Lakoff and Mark Johnson, regarding the prominent role that metaphors play in cognition (1980; 1999). By applying their general account to the particular case of introspection, I will suggest that the enduring attractiveness of the perceptual account of introspection can be explained by the prominence of perceptual metaphors in human cognition. People regularly use the idea of perception to understand other sorts of phenomena. In particular, we project the concrete, rudimentary familiarity we have with perception

as a metaphorical framework for how knowledge works, and this is why we tend to understand self-knowledge in a perceptual manner.

Following the "embodiment" trend in cognitive science, Lakoff and Johnson suggest that our higher cognitive abilities originate out of bodily sensorimotor processes. Roughly, the idea is that primary body functions, such as vision and movement, form the basis of our concepts and ways of thinking about how the world works. In turn, these basic conceptual structures are abstracted and applied metaphorically to other domains, which allows us to develop higher-level cognitive abilities. Here is how Lakoff and Johnson describe it:

> Our subjective mental life is enormous in scope and richness. We make subjective judgements about such abstract things as importance, similarity, difficulty, and morality, and we have subjective experiences of desire, affection, intimacy, and achievement. Yet, as rich as these experiences are, much of the way we conceptualize them, reason about them, and visualize them comes from other domains of experience. These other domains are mostly sensorimotor domains, as when we conceptualize understanding an idea (subjective experience) in terms of grasping an object (sensorimotor experience) and failing to understand an idea as having it go right by us or over our heads. The cognitive mechanism for such conceptualizations is conceptual metaphor, which allows us to use the physical logic of grasping to reason about understanding. (1999, p. 45)

So, from this perspective, metaphors are much more than literary devices. They serve as basic conceptual structures that play a constitutive role in human cognition. According to Lakoff and Johnson, this cognitive use of metaphor is pervasive, encompassing the full range of human experience and thought, from emotions and relationships to epistemological theories and lofty metaphysical speculations. I do not wish to make any commitments here regarding the extent to which Lakoff and

Johnson's perspective can account for human cognition in general. Rather, my purpose is to apply it to the much more limited topic of introspection. Whether or not there are other domains of thought that utilize embodied metaphorical projections, the understanding of introspection in terms of perception clearly involves this type of cognition.

In order to *see* this, let us first *look* at the metaphorical use of perception to understand knowledge more generally. Lakoff and Johnson note that

> We get most of our knowledge through vision. This most common of everyday experiences leads us to conceptualize knowing as seeing. Similarly, other concepts related to knowing are conceptualized in terms of corresponding concepts related to seeing. In general, we take an important part of our logic of knowledge from our logic of vision. Here is the mapping that projects our logic of vision onto our logic of knowledge.
>
> The Mind Is A Body
> Thinking Is Perceiving
> Ideas Are Things Perceived
> Knowing is Seeing
> Communicating is Showing
> Attempting to Gain Knowledge Is Searching
> Becoming Aware Is Noticing
> An Aid to Knowing Is A Light Source
> Being Able to Know Is Being Able To See
> Being Ignorant Is Being Unable To See
> Impediments To Knowledge Are Impediments To Vision
> Deception Is Purposefully Impeding Vision
> Knowing From A "Perspective" Is Seeing From A Point Of View
> Explaining In Detail Is Drawing A Picture
> Directing Attention Is Pointing
> Paying Attention Is Looking At
> Being Receptive Is Hearing
> Taking Seriously Is Listening
> Sensing Is Smelling
> Emotional Reaction Is Feeling
> Personal Preference Is Taste (1999, p. 238)

Again, I do not want to commit myself to the full extent of Lakoff and Johnson's theory. Arguably, there is more to knowledge than just vision and the projection of perceptual metaphors. But, it is clearly one way in which knowledge gets conceptualized. We can all think of particular instances of the above metaphorical paradigms, in our own thoughts and words as well as others. I, for instance, have repeatedly written in this dissertation of "looking" at arguments and considering "viewpoints" or "perspectives." I am sure we have all said and heard the words "I see what you mean" innumerable times, without giving even the slightest notice to the metaphorical nature of the statement. So, it is clear that such language is a prominent way of conceptualizing knowledge and understanding. Yet, none of these sorts of conceptualizations can be said to be what is literally going on in our minds. We don't actually see meanings, and philosophical theories are not viewpoints in the literal sense of perception.

I suggest that this sort of metaphorical conceptualization is what is really going on when people think of introspection in terms of inner perception. We do not actually see or otherwise perceive the diverse aspects of our mental lives. We do not literally look within ourselves in our more self-reflective moments. These are just ways of conceptualizing self-knowledge, in terms of something that we are already familiar with. Getting an actual grasp of what literally goes on in our minds when we introspect is really quite difficult, and can lead us into confusing and seemingly paradoxical territory (*see*, for instance, Hofstadter 1979 and Bermúdez 1998). So, it should really be no surprise that we use our intuitive understanding of perception as a

metaphorical projection to get a grasp on what happens in introspection. However, as I have shown in the preceding section of this chapter, it would be a mistake to take this understanding as a literal account. If we want to understand the ways in which we actually acquire self-knowledge, we need to set the perceptual metaphor aside.

Following this reasoning, some might be tempted towards an eliminativist stance on introspection (Lyons 1986 seems to head in this direction, for instance). If there is no internal perception involved in introspection, why should we even keep the term? After all, it is derived from the Latin words for "to look" and "within" (*spicere* and *intra*, respectively) (Lyons 1986, p. 1). I want to shy away from this eliminativist perspective, however. Although introspection cannot be literally understood as inner perception, the term and the metaphorical concept can play perfectly legitimate roles in the right contexts. It is already in place in our common understanding, just as the idea of a sunrise is a persistent metaphor. The sun does not literally rise, of course, but we still find it useful to conceive of it in that way for everyday purposes. Perhaps something similar can be said for introspection. As Lakoff and Johnson suggest, metaphors play a significant role in human cognition. Along these lines, it is perfectly normal, in terms of actual human cognition, to use the perceptual metaphor to understand knowledge of our own mental states. However, it is important to acknowledge this metaphorical understanding as such, and not regard it as a literal account of what goes on in a person's mind.

I am even willing to go so far as to entertain the possibility that there is some kind of virtual perceptive process in the brain that is created by the metaphorical use

of perception. Perhaps, roughly following Dennett's conception of the mind, introspection is a somehow real but simultaneously fictitious construction of the brain, in the sense that our minds implement a kind of introspection "program" as software in the mind's functionally plastic hardware (Dennett 1991; 1998, esp. pp. 95-120). As I will discuss later, there is a sense in which our minds create themselves, so perhaps the mind could create for itself a purely functional device that takes its own states as input (See Hofstadter 1979 for interesting explorations of this idea). Even if this is a possibility, though, it does not follow that we literally perceive our minds, in the same way that we perceive aspects of our external environment. Rather, this sort of introspection would count only as a simulation of perception. It would be like a Turing machine can opener, and not like one in your kitchen that can actually open cans.

So, in conclusion, the understanding of introspection as a kind of perception is only a way of conceptualizing how our minds work. It is a commonsensical metaphor that gives us a way of talking about ourselves, but not a literal account of what goes on in our minds. Just as we talk about the sun rising and setting, we talk about "looking within" or "seeing ourselves in a new light." Equally, though, we know that it is not the sun moving but rather the earth that's turning; and we now know that we do not perceive our own minds.

CHAPTER 3 – THE DIVERSITY OF INTROSPECTION

So if introspection is not a kind of perception, then what is it? To provide an accurate account of introspection, it is crucial to recognize that it consists of a heterogeneous collection of processes, rather than a single unitary process. Philosophers and scientists often have a tendency towards unifying whatever phenomena they are dealing with, either by drawing generalized patterns from data or constructing necessary and sufficient conditions to explain some concept, fact, or process. Of course, this tendency is based on processes that we need in order to understand ourselves and the world we live in, but it can also cause us to overlook the inherent complexity of some phenomena. I contend that the latter is the case with introspection. There is no single cognitive phenomenon we can point to and say "that's introspection." Instead, as we saw at the end of the last chapter, the idea of introspection is a metaphorical projection of perception onto the diverse sorts of things that we can know about ourselves. The unifying feature of introspection is not a particular process but rather the focal point on oneself as an object of knowledge. Introspection is the acquisition of self-knowledge. As such, introspection consists of processes through which we obtain information and understanding about ourselves from our own first-person perspective. A diverse assortment of states and processes falls under this category:

1) the phenomenal awareness involved in one's own conscious states, such as an immediate feeling of a pain

2) attention processes involved in one's own experience and cognition, for instance attending to some particular auditory sensation, such as a linguistic utterance, or remembering some past event in one's own life

3) representation and theory-like conceptualization of one's own mental states, most prominently involving the attribution of beliefs and desires to oneself

4) construction of linguistic / propositional structures about oneself, including rationalization and narration about one's own actions and life events

We obtain information about ourselves through this heterogeneous collection of psychological processes, which vary significantly in their epistemic qualities. It is important to emphasize this epistemic diversity. All too often, as seen in the views dismissed in Chapter 2, introspection is mischaracterized as a single kind of process, and with this comes the erroneous idea that there is a single epistemic dimension to introspection. Some people regard it as infallible or epistemically privileged in some way, while others denigrate introspection as significantly biased, untrustworthy, and subjective. However, whether it is dogmatic insistence upon infallibility or the cursory dismissal of introspection often found in the beginning of psychology textbooks, all such unitary assessments of introspection are misguided. We cannot characterize the nature and epistemic qualities of introspection under a single category or formulaic expression. Introspection as a whole is neither infallible nor

epistemically deplorable. Introspective information comes from a heterogeneous collection of processes, each with their own unique epistemic pros and cons. So, we have to consider which domain of introspection we are talking about if we are to say something useful or accurate about it. This is what I propose to do in the following chapters. I will illustrate and analyze importantly different modes of introspection, in attempt to better understand them and assess their unique epistemic qualities. In doing so, I will address each of the 4 different modes of introspection listed above.

CHAPTER 4 – KNOWING OUR OWN CONSCIOUS STATES

4.1 Introduction

I will begin this excursion into the varieties of introspection by discussing the immediate phenomenal awareness involved in our conscious experiences. A number of philosophers have brought attention to the qualitative, experiential aspect of consciousness, and I will draw on this attention to explicate an immediate kind of knowledge that we have in virtue of *being constituted by* such conscious states. As a kind of self-knowledge, the phenomenal knowledge of one's own conscious states can be seen as a rudimentary sort of introspection, with unique epistemic properties. I hope to characterize this kind of knowledge in a way that falls between the cracks of the more contentious issues regarding the metaphysical status of these states, which is typically the focal point in discussions of consciousness. In fact, along the way, I will suggest that conscious experiences in themselves tell us nothing about their metaphysical status. So, rather than taking a stand on the metaphysical nature of consciousness, my purpose is to bring to light the epistemic aspects of phenomenal conscious experiences, particularly in regard to the knowledge that experiencing agents obtain of themselves through such experiences. I will begin by introducing phenomenal conscious states as they have been characterized through the work of Thomas Nagel and Frank Jackson. I will then explain the kind of knowledge that such states create, and defend the idea that this constitutes a rudimentary kind of

introspective knowledge. Finally, I will offer some clarifications regarding the nature of this kind of self-knowledge, explaining how general skepticism towards self-knowledge erroneously ignores the unique epistemic characteristics of this particular kind of introspective knowledge.

4.2 Knowing What It Is Like

To get a grasp on the kind of self-knowledge that is involved in the immediate phenomenal awareness of one's own conscious states, I will overview two paradigmatic illustrations given by the philosophers Thomas Nagel and Frank Jackson, both of which suggest that there is something special about conscious states in virtue of their subjective, phenomenal aspects.

In his highly influential article "What is it like to be a bat?" Nagel draws attention to the subjective, felt quality of conscious mental states (1974). The use of the phrase "what it is like" has become a common way of referring to the qualitative aspect of conscious experience, so it is worth looking at what exactly Nagel had in mind when he brought the phrase into frequent usage among philosophers. He states, "an organism has conscious mental states if and only if there is something that it is like to *be* that organism – something it is like *for* the organism" (1974, p. 519). So, conscious states, as Nagel characterizes them, are constituted by experiences that an organism has, as a subject. There is some phenomenal quality to the organism's experience, such that there is something it is like for the organism to have the

experience, or to be in that particular state. Notice also the fact that Nagel puts the term "*be*" in italics, drawing attention to the existential dimension of conscious experience. Although Nagel does not put it this way, a crucial feature of conscious experience is the fact that organisms are, at least in part, constituted by them. This will be important later when I characterize the kind of knowledge that such conscious states engender.

In conjunction with his characterization of consciousness, Nagel suggests that consciousness fails to be encapsulated by objective, physical knowledge. Indeed, Nagel's primary point is to illustrate the inability of physical knowledge to account for conscious experience. To this effect, he suggests that no amount of knowledge concerning the physiology of a bat can tell us what it is like, on the inside (so to speak), for a bat to experience echolocation. Only a being that undergoes the experience of echolocation can know what it is like to have such an experience. Objective physical information simply cannot portray to us the subjective, experiential dimension of conscious states, or so it seems to Nagel. From this basis, Nagel concludes that consciousness cannot be accounted for from a purely physicalist perspective (at least not without radical changes in our understanding of objective, physical reality). However, I want to separate myself from this further claim. In fact, I will later offer a criticism of such inferences. For the time being, however, I want to focus only on the qualitative, experiential character of consciousness. Irrespective of whether or not they can be accounted for in physical terms, conscious experiences are constituted by phenomenal qualities such that there is something it is like to have, or

be constituted by, them. Regarding this portrayal of conscious mental states, I think Nagel is right on track. There are qualitative aspects to feeling pains, seeing colors, hearing sounds, having orgasms, and so on, such that we can say there is something it is like to be in such states.

Although he explicitly separates his arguments from Nagel's, Frank Jackson offers some similar concerns regarding the unique experiential character of consciousness. To illustrate his position, Jackson proposes the hypothetical example of a fully informed neuroscientist, Mary (1982, 1986). Mary knows all the physical facts regarding the neurophysiology of vision, understanding completely and exactly what goes on in the brain when people see. Of course, Mary must be someone in the future, as we are currently far from a complete understanding of vision, but let us set all practical concerns aside and assume for the sake of argument that the neuroscience of Mary's time is fully accurate and comprehensive. Now, there is one important catch about Mary's situation: She has been confined to a black and white environment from birth. All of her surroundings are cast in black, white, and varying shades of gray, and she has somehow been prevented from seeing the various colors of her body, presumably including even the reddish hues seen through closed eyelids in a brightly-lit environment (How this could be achieved, I have no idea. But, again, we can set aside the practicalities of Mary's circumstances). Given this scenario, Jackson goes on to consider what might happen if Mary is suddenly allowed to experience color:

> What will happen when Mary is released from her black and white room or is given a color television monitor? Will she *learn* anything

> or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had *all* the physical information. *Ergo* there is more to have than that, and "physicalism" is false. (1982, p. 558)

Like Nagel, Jackson's primary motivation is to show that the phenomenal aspect of experience escapes physical understanding. When Mary finally comes to experience colors, she purportedly learns something that no amount of physical information could possibly give her. There is some kind of new knowledge acquired when she first experiences color vision, and because of this Jackson concludes that there must be more to phenomenal states than physical processes. However, as previously mentioned, I want to separate myself from claims about the relationship between phenomenal mental states and physical states. I think that Jackson is much too quick to conclude that phenomenal states must have something non-physical about them. This will be touched on later, but for now I simply want to draw attention to the fact that Mary does learn something new when she experiences color vision for the first time. Although Mary thoroughly understood how color vision happens, she did not know *what it is like* to actually undergo color vision when she was confined to her black and white environment. Only by actually having a color vision experience does Mary come to know the qualitative, phenomenal aspect of consciously seeing, say, the particular red hue of a fire engine or the drab green color of an olive. Although I disagree with his conclusion, I think that Jackson's thought experiment here is quite useful in drawing attention to such unique phenomenal characteristics of conscious experiences. There is indeed something it is like to the

particular conscious experiences involved in color vision, such that one can only know these experiences by actually having them. This constitutes a very basic kind of self-knowledge, in that one knows what one's own conscious experiences are like simply by, and only by, having them.

4.3 What Kind of Knowledge Is This?

If we obtain knowledge through the qualitative phenomena of conscious experience, then what kind of knowledge is this? Following the wake of Nagel and Jackson, a number of viewpoints have been expressed on this topic (for a recent overview, see Nida-Rümelin 2002). It would take us astray to consider all of the options here, so I will only take a look at the view that I think has come closest to getting it right and then move on to articulate and defend my own view on the topic. In brief, I will suggest that the knowledge involved in phenomenal awareness is a kind of non-propositional knowledge that we have by being (at least in part) constituted by conscious states. In other words, it is not knowledge *that* such and such is the case, but rather knowledge *of* our own conscious states, acquired by *being in* those states.

In a response to Jackson, Earl Conee suggests that when Mary experiences color for the first time, she does not acquire knowledge of any new, mysteriously non-physical fact (1994). Instead, what she acquires is acquaintance with a phenomenal experience that she has for the first time. To this effect, Conee states:

> A simple acquaintance hypothesis about what Mary learns is that learning what an experience is like is identical to becoming acquainted with the experience. When Mary first sees red ripe tomatoes, she learns what it is like to see something red. It is also true of this episode that it is the first time that she undergoes an experience with the phenomenal quality that ordinarily results from seeing something red, phenomenal redness. Suppose that this quality is a physical property of experiences. If phenomenal redness is a physical property, then from the assumption that Mary already knew all of the physical facts it follows that she already knew that experiences have this property. But during her confinement phenomenal redness was not a property of any of her visual experiences. It seems also to be true that she never knew the property itself, in spite of her knowing all about it. This suggests the more specific acquaintance hypothesis that becoming acquainted with a phenomenal quality consists in experiencing the quality. This further hypothesis puts us in a position to account for Mary's learning what it is like to see something red. The learning is a matter of Mary's becoming acquainted with the visual experience that ordinarily results from seeing something red, and this acquaintance consists in Mary's experiencing phenomenal redness. She experiences the quality, and that teaches her what seeing red things is like. She does not learn any new fact. Rather, she comes to know the quality itself. (1994, pp. 140-141)

So, on this basis, Conee argues for the presence of a special kind of knowledge, namely acquaintance knowledge, that is involved in knowing what conscious states are like (for an earlier characterization of acquaintance knowledge, see Russell 1910). In coming to know what experiencing red is like, Mary is acquainted with the phenomenal quality of redness, much as we come to know another person by meeting her or come to know a city by spending time in it. In all such cases, the basis of knowledge is an acquaintance relationship between the knower and the thing known. It is crucial to note that this kind of knowledge is fundamentally different than propositional knowledge. In all cases of acquaintance knowledge, there is no particular propositional fact that is known. In Mary's case, she knows all of the facts

already. When she comes to know what it is like to see red, there is no further proposition X such that Mary learns for the first time *that* X is the case. Instead, she becomes acquainted with the experience of red in a way that only a person that actually experiences red can, and that is the integral feature of her coming to know what it is like to see red.

I think that this non-propositional approach to understanding phenomenal knowledge heads us in the right direction. Knowing what it is like to have a particular kind of conscious experience does not involve assent to a claim about such experiences. For instance, in knowing what it is like to have an orgasm (just to provide some contrast to the fixation on pain that seems to pervade the philosophy of mind), there are not any claims that a person must accept or deny, either explicitly or tacitly. It simply does not involve any propositional content. Of course, there can also be propositional content involved in one's understanding of a phenomenal state, such as "Having an orgasm is more pleasurable than having a toe cut off," but any such proposition is tangential to the actual experience of a phenomenal state, which is what forms the content of knowing what a particular conscious experience is like. So, in this respect Conee's acquaintance account of phenomenal knowledge accurately reflects the nature of such knowledge.

Moreover, the realization that phenomenal knowledge is non-propositional shows us where Nagel and Jackson go wrong in drawing metaphysical conclusions on the basis of phenomenal conscious experience. The experience involved in a conscious state does not in itself provide us with any factual knowledge about that

state. An experience does not tell the experiencing agent *that* it is a physical event or a non-physical event. It is simply silent, so to speak, on the matter. To make any such judgment, the agent must engage in reflective, propositional thought about the experience, which requires theory-like conceptualization (more on this later). If the agent is operating under certain assumptions about the nature of physical reality, then he may conclude that his experience cannot be a physical event. Alternatively, however, he may have a different understanding of things that lead to the conclusion that the experience is a physical event. Either way, the agent is going beyond the experience itself in making any such judgment about it. Now, I'm willing to grant Nagel and Jackson the point that, given our current understanding of physical reality, it is indeed puzzling how conscious experiences can come about through physical processes. However, it is just as likely (if not more likely) that this puzzlement is due to problems in our understanding of physicality as it is that consciousness is a non-physical event. Conscious experiences in themselves, however mysterious they may seem, simply do not preclude the possibility that they are physical events. Perhaps that is just what physical reality is like, when known from the unique perspective of being a particular kind of physical event. As Bertrand Russell once said, "we know nothing about the intrinsic quality of physical events except when these are mental events that we directly experience" (1956, p. 153). By drawing attention to the non-propositional nature of phenomenal knowledge, the acquaintance account helps us see this as at least a possibility (and for the present point, that is all I'm saying that it is). Phenomenal knowledge in itself tells us nothing about metaphysics.

However, I find it somewhat misleading to characterize this kind of knowledge as acquaintance knowledge. Knowing something through acquaintance implies that there is some kind of relationship or interaction between an epistemic agent and the known object. Consider, for instance, Russell's view that we know sense-data through acquaintance. He states "When we ask what are the kinds of objects with which we are acquainted, the first and most obvious example is *sense-data*. When I see a colour or hear a noise, I have a direct acquaintance with the colour or the noise." (1910, p. 109). Similarly, Conee suggests that phenomenal knowledge consists of being acquainted with phenomenal qualities, such as redness (1994, pp. 140-141). Now, one immediate problem with these views is that it is contentious whether or not sense-data or phenomenal qualities even exist as objects to be known. The idea of sense-data has been largely dismissed among analytic philosophers, and if we construe phenomenal qualities like redness as objects to be known within the mind, then they arguably ought to suffer a similar fate. As I argued in Chapter 2, we do not literally observe any objects in our own minds. But, even if one could construe the acquaintance relationship in some way that avoids the problematic idea that we observe mental states, I still think that it would be misguided. The kind of knowledge involved in knowing what a phenomenal state is like is not relational at all. There is no dichotomy here between the one who is acquainted and that with which he is acquainted. Instead, the knower knows what his own phenomenal states are like because he is, at least in part, composed of those states! Consider again the example of an orgasm. When you experience an orgasm,

and thereby know what it is like for you to experience an orgasm, you have this knowledge in virtue of *being* the one having the orgasm. At that moment, *you* are partially *made up of* an orgasmic state. In terms of your experience in itself, there is no separate orgasmic object with which you are acquainted. Rather, you yourself are in an orgasmic state. So, along these lines, I think it is wrong to think of knowing what it is like as a kind of acquaintance knowledge. Perhaps a better label would be "existence knowledge", in that it is knowledge we have of ourselves in virtue of being ourselves, or "constitutional knowledge", in that it is knowledge one has in virtue of being constituted by a particular state. However, I see no reason why we cannot simply continue to use the colloquial phrase "knowing what it is like" to refer to this epistemically unique kind of knowledge. Regardless of the label, the key point to grasp is that we know what phenomenal conscious states are like in virtue of the fact that we, whatever else we may be, consist of such states. In regard to the immediate knowledge of conscious experience, we do not know of phenomenal states as objects with which we are acquainted, nor as facts that we grasp propositionally. Rather, we know them by *being* them.

This is not to say that we cannot *also* know phenomenal states as objects. In fact, as I will explain later, we do need to objectify or conceptualize our experiences in some manner in order for us to think about or cognitively reflect upon them. Clearly, we human beings have the capacity to obtain knowledge of our own mental states by forming and applying concepts, treating our experiences as objects to be known in propositional terms. But, this kind of self-knowledge extends beyond the

purely experiential aspects of self-knowledge that I am presently illustrating. The present point is only that our conscious experiences in themselves should not be thought of in this objectifying way. We do not know a conscious experience, qua the experience in itself, as an object, but only as the subjective experience or feeling of being that state. There is no dichotomy between a knower and a known object here, but only after some kind of cognitive abstraction takes place.

4.4 Objections and Clarifications

I find the idea that knowing what a conscious state is like constitutes a unique kind of epistemic state to be both intuitively appealing and explanatorily useful. It makes sense of the fact that we have qualitative, phenomenal experiences that seem ineffable through any other means besides actually having them, and helps us understand what makes this experiential kind of knowledge epistemically unique. However, I anticipate that others will not find this perspective as satisfying as I do. So, I will now address some potential objections to the idea that knowing what it is like provides a unique kind of introspective knowledge. In doing so, I will also offer some clarifications that will further elaborate upon my view, hopefully making it more appealing and understandable.

One possible concern is that knowing what it is like is not really a kind of knowledge at all. The standard viewpoint among epistemologists is that knowledge consists of justified true beliefs. However, if my view of phenomenal knowledge is

correct, such knowledge has no propositional content. Consequently, it could not possibly consist of true (or false) beliefs, and it is unclear as to what it would even mean for a phenomenal state in itself to be justified or unjustified. For instance, the experience of an orgasm in itself is neither true nor false, and determining whether or not it is justified is, if anything, more an ethical issue than an epistemological one. So, in what sense could such a state be considered an instance of knowledge? I think that this issue is only definitional, and that we merely need to recognize that there are different senses of knowledge. If we define knowledge as justified true belief, then it does indeed seem to exclude the kind of knowledge that I have portrayed. Yet, it seems perfectly natural for us to speak of knowing what phenomenal states are like. If someone were to ask you "Do you know what the taste of salt is like?" amidst a conversation about food, I anticipate that you would say "yes" without any hesitation or confusion about the meaning of the question, unless, of course, you have never tasted salt or you are a philosopher in the grasp of some theory about knowledge (but even then I think you'd be likely to slip up and say "yes," if the question occurred in some ordinary, non-reflective context). People talk about knowing experiences, in the phenomenal, qualitative sense that I have been describing, all the time. So, on this basis, we can at least say that there is some sense in which knowing what experiential states are like is a kind of knowledge. Knowing what an experience is like gives us some kind of knowledge, in that we become cognizant of something about ourselves, or the world, or perhaps even both at the same time. We obtain some kind of grasp on the nature of things that we did not have before, prior to the

experience that is known in this manner. So, by simply broadening the scope of knowledge beyond the strict sense of justified true belief, we make room for saying that we know what phenomenal states are like.

Even if we do not count knowing what something is like as a genuine instance of knowledge, however, it still serves a clear epistemic purpose. Phenomenal conscious states are not themselves beliefs, but they do play integral roles in our belief systems. Whether we are operating from a foundationalist, coherentist, or some other epistemological framework, beliefs and experiences are intimately related. At a very minimum, many of our beliefs are about our experiences. For instance, your beliefs regarding the tastes of foods, such as what foods taste similar or dissimilar, what foods you like and dislike, and so on, are all tied to your phenomenally conscious experiences with foods. Going beyond this, from a broadly foundationalist perspective, we could add that your experiences provide justification for your beliefs, such as when you form the belief that some dish is spicy on the basis of a "hot" sensation in your mouth. Of course, there is much more to say about the relationship between experiences and beliefs, but for the present point all we need to take note of is the fact that there is *some* epistemic relation between them. So, the phenomenal experience involved in the kinds of conscious states I have been discussing has an epistemic component. I prefer to characterize this under the colloquial category of "knowing what X is like", but even if you are hesitant to regard this as a proper kind of knowledge, there is still room to regard phenomenal experience in epistemic terms, in that it underlies or relates to belief systems in some manner.

Another potential objection to my view is that phenomenal experience is not introspective. Introspection is often regarded as being self-reflective in some way. For instance, as we saw in Chapter 2, perhaps the most common understanding of introspection is that it involves "looking within" oneself. So, all worries about the literal accuracy of this understanding aside, one might plausibly think that introspection, whatever else it may be, is inherently self-directed. I'm not introspecting unless I'm explicitly directing my thoughts inward, towards myself. If so, then it may seem as though introspection excludes phenomenal experiences in themselves. A particular experience of red, for instance, need not be self-reflective in any way. When Mary sees a red fire truck for the first time, she may not be thinking about herself at all. So, how then could this possibly count as a kind of introspection?

There are a couple of things that need to be said here. First of all, I grant that there are numerous ways of introspecting that involve going beyond phenomenal experiences in some way. In fact, in the chapters that follow, I will describe several different modes of introspection that are clearly and explicitly more self-reflective than phenomenal experience. Again, I think that introspection is a heterogeneous phenomenon, which does not admit of any singular explanation. So, we should not be surprised if any one kind of introspection does not match up with all of our intuitions about what introspection involves. Admittedly, phenomenal states are not introspective in the sense that introspection involves self-reflective thought. Nevertheless, there is still a sense in which we can legitimately regard phenomenal states as providing a kind of rudimentary introspection. Phenomenal experience, in

the sense that it involves knowing what something is like, provides some intrinsic understanding of the experiencing organism's own conscious states. Consider again Nagel's statement that "an organism has conscious mental states if and only if there is something that it is like to *be* that organism – something it is like *for* the organism" (1974, p. 519). The feature of being something it is like *for* the organism is crucial here. While conscious experience in itself is not self-reflective, it is intrinsically self-regarding in virtue of this very basic, and easily overlooked, aspect. When an organism knows what an experience is like, it thereby automatically knows something about itself, namely, what its experiences are like. It knows itself in the most basic and intimate way possible, in virtue of being itself. So, in light of this feature of conscious experience, I think we can legitimately say that phenomenal experiences have an intrinsically introspective element to them. While Mary may not be explicitly entertaining thoughts about herself when she sees red for the first time, the experience itself gives her knowledge of herself. She knows what her own phenomenal experience of seeing red is like.

José Luis Bermúdez has offered some related points that will be helpful to consider here. According to Bermúdez, both perception and proprioception contain inherently self-specifying content (1998, pp. 103-162). Following J. J. Gibson's theory of ecological optics, Bermúdez illustrates in detail how an organism tacitly obtains information about itself through the constantly shifting perceptual array it encounters while moving about in an environment. For instance, when you walk from one room to another, the perceptual changes you visually encounter inherently

specify to you your location in relation to the objects around you. Similarly, the various modes of proprioception (detection of muscle positions, balance detection through the vestibular system in the inner ear, etc.) provide inherently self-specifying information to organisms that have them. For example, our limbs contain various nerve receptors that indicate their positions to us. I will take up proprioception again in the next chapter. The important point here is that, by its very nature, proprioception provides information with self-specifying content. In recognition of this, Bermúdez regards proprioception, along with visual perception, as a primitive form of self-consciousness, detailing the various ways in which it provides an organism with information about itself (1998, pp. 131-162).

To be clear, I am not concerned with the informational content of these states here. Instead, the key point is that the content is self-specifying. I mention this because I think that phenomenal states have a similar self-specifying nature. The experience of an orgasm, for example, provides immediate and intimate knowledge about the experiencing agent to that same agent. By its very nature, such a phenomenal state confers self-understanding in the most primitive manner possible. As discussed earlier, it is crucial to recognize that knowing what a phenomenal experience is like intrinsically involves *being constituted by* the experience itself. The experiencing agent of a phenomenal state is itself made up of that state. So, in this respect, knowing what it is like is an intrinsically self-specifying form of knowledge. While this is clearly not the complicated sort of self-reflection that we

often have in mind when we speak of introspection, it is nonetheless a very basic kind of self-understanding.

Before moving on, I want to consider one other possible concern. In Chapter 2, I criticized the perceptual account of introspection on the basis that there is no identifiable organ that literally perceives mental states. However, one might think that a similar criticism could be lodged against the view I am defending in this chapter. I have not said anything about a physical or empirically identifiable basis for conscious experience, so how then can I speak of it with such assurance? In response to this concern, it is important to recognize the difference between a phenomenon to be explained and an explanation of that phenomenon. The view that introspection is a kind of perception is an explanatory account of what introspection is and how it works. It proposes a specific mechanism through which introspection purportedly occurs, and in respect to that I urged that some such mechanism needs to be actually identified if it is to be regarded as a legitimate account of introspection. Phenomenal conscious experiences, on the other hand, are not explanations in themselves but rather phenomena to be explained. So, no specific underlying mechanism or account is necessary to discuss them. How and why conscious experiences occur is a fascinatingly difficult issue, but conceptually speaking, it is tangential to the fact that they occur and provide intrinsic self-knowledge to agents that have them. Whatever they may be, conscious experiences do occur and they constitute a kind of self-specifying knowledge through their very occurrence. For my purposes here, this is all we need.

4.5 Putting Skepticism of Self-Knowledge in its Place

Now that I have explained and defended my view, I want to use it to make some important clarifications regarding the epistemic status of introspection. Of course, discerning the epistemic properties of introspection is a running theme throughout this work, but here I am particularly concerned with a kind of skepticism that has been directed towards introspection in general. While there are some grounds for skepticism regarding certain domains of self-knowledge, the broad sense of skepticism that I will address here erroneously ignores the unique epistemic features of the kind of phenomenal self-knowledge I have been describing in this chapter. I will identify and correct this error by discussing it in a place where it is particularly evident: Hilary Kornblith's skepticism of self-knowledge (1998, 2002).

In his "What Is It Like to be Me?" Kornblith presents a skeptical view of self-knowledge, particularly in regard to the role of introspection in obtaining knowledge of one's own mental states (1998). I believe that this view misconstrues the nature of knowing what one's own conscious experiences are like, thus leading to an over-generalized skepticism about self-knowledge. As I have stated already, there are importantly different types of introspection, which have correspondingly different epistemic qualities. So, the prospects for self-knowledge will vary in accordance with the type of introspection in question. Kornblith, however, does not acknowledge any such qualification, and this leads to his over-stated skepticism.

Obviously, Kornblith is responding to Thomas Nagel's article "What Is It Like to Be a Bat?" (1974). However, his use of the "knowing what it is like" terminology, as the title "What Is It Like to be Me?" suggests, is rather different than Nagel's. In fact, I believe that Kornblith misuses the phrase, failing to acknowledge Nagel's purpose in applying the phrase to philosophical concerns in the first place. As explained earlier, Nagel wants to draw attention to the subjective, experiential nature of consciousness. However, it is precisely this experiential character of conscious mental states that Kornblith leaves out in his use of "what it is like," and which leads to his wrongly focused proposal that a person can be wrong or ignorant about what it is like to be himself or herself. To explain how this is the case, I will use one of the examples Kornblith discusses.

He presents us with a man, Jack, who lacks an accurate understanding of his relationships to others and of his own character traits and emotions. For instance, Kornblith states that

> Jack is very defensive. He is quite insecure, and he believes, incorrectly, that people are frequently talking down to him. … Jack does not believe that he is defensive, insecure, or self-involved. Jack's unreflective opinion of himself is quite different; it is terribly inaccurate. … Jack's understanding of his emotions is no more accurate than his understanding of his character. I have already said that Jack becomes angry and defensive when he is criticized in any way. But Jack sincerely believes that he does not become angry when criticized… (1998, pp. 50-51)

Obviously, Jack is lacking in self-knowledge. Given this description, there is no question that his understanding of himself is quite wrong, and in multiple ways. From this basis, Kornblith goes on to ask "What is it like to be Jack?" and concludes

that Jack himself would fail to know the answer. Jack does not know what it is like to be Jack because his introspective capacities are badly askew.

This conclusion is rather misguided. Of course Jack is wrong about his emotions and character traits, but these are not facts about what it is *lik*e to *be Jack*. What it is like to be Jack is what it is like to be someone who experiences his particular emotions and character traits (including the experience of misunderstanding them), irrespective of whatever beliefs may be constructed about the experiences. So, knowing what it is like to be Jack is not the same as knowing what kind of person he is. Rather, it is knowing what the actual experience of such a person is like, from the "inside," so to speak. There is a crucial difference in kinds of self-knowledge here, to which Kornblith does not give adequate attention.

Kornblith verbally acknowledges that there is something it is like to be Jack, but denies that such experience, in itself, constitutes any kind of self-knowledge. In order to have self-knowledge, Kornblith holds that one must be able to conceptualize such experience and form beliefs about it. To this effect, he states that "the experience of being oneself, without any belief about that experience, is no kind of knowledge at all" (1998, p. 53). This perspective follows the presumption that knowledge is "knowing that." Knowledge, in this sense, requires assent to some belief, or to some proposition, which one knows. If, however, one accepts this restricted definition of knowledge, then it would seem that one must also refrain from talk of "*knowing* what it is like," as this terminology inherently implies the excluded sense of knowledge. Kornblith, as we have already seen, does not keep to this requirement. By discussing

"knowing what it is like" and simultaneously claiming that experience is not a type of knowledge, Kornblith is making a tacit but fundamental contradiction in his overall account. This creates confusion in the attempt to analyze one's ability to "know what it is like," as it literally entails that "what it is like" is no type of knowledge at all. If knowledge requires propositional beliefs, then it simply does not make sense to discuss "knowing what it is like."

As explained earlier, I think it makes sense to discuss "knowing what it is like." It is important to point out here that such knowledge consists in the very experience itself. The knowledge exists prior to any proposition that may be given about it. On this basis, I maintain that knowing what it is like for a person to *be* himself or herself is a real and immediate source of self-knowledge. I am not saying here that this type of self-knowledge is exhaustive, nor that it constitutes a large portion of the many facets involved in understanding ourselves. Rather, I am saying that our conscious experience of ourselves is *one* thing that we do know about ourselves. It is something we know intimately and directly, by the very occurrence of our experience, prior to anything that may be stated about it. This is unique to the "what it is like" perspective and is an important factor in giving an accurate account of self-knowledge. To ignore this uniqueness, as Kornblith does, leads to an inaccurate picture of self-knowledge.

A related point that needs attention here is Kornblith's positive account of how one can know "what it is like." Kornblith claims that Jack would not be a good source for learning what it is like to be Jack. Kornblith proposes, rather, that to know

what it is like to be Jack, "We will need to engage in theory construction. We will need to try to figure out, given his behaviour, including his verbal behaviour, what the mental life of such a person must be like" (1998, p. 51). Underlying this proposal is the assumption that there is no qualitative difference between first-person and third-person knowledge. To this effect, Kornblith states that

> Because the difference between a first-person perspective and a third-person perspective is a matter of causal proximity, it is comparable to the difference between looking at a person from across a table and from across a street: closer is, frequently, better. (p. 55)

So, according to Kornblith, third-person objective knowledge can potentially tell us (as long as it can get "close enough") all there is to know about what it is like to be a given person, such as Jack. Such knowledge would not be different in kind from that which is obtained from a first-person perspective, by the person in question.

In complete contrast to Kornblith, I believe that Jack, despite the fact that he would be a terrible source of information regarding his own character traits, would be a good source in regard to what it is like to *be* him (setting aside worries about his reports of this knowledge). In fact, he is the *only* source of this particular kind of knowledge because he is the only epistemic agent that actually has, or consists of, the experiential states in question. Within the context of experiential states, the first-person, or subjective, perspective is different in kind from the third-person, or objective, perspective. These are qualitatively different epistemic conditions. The first-person, subjective perspective is characterized by knowing *what* X is like, while the third-person, objective perspective is characterized by knowing *that* X is Y. The former kind of knowledge is intrinsic to the knowing subject, while the latter is

essentially relational, involving the knowing subject's conceptual relationship to the known object. By denying this qualitative distinction and assuming that it is conceivable to give a third-person account of first-person experience, Kornblith is simply ignoring the unique epistemic nature of knowing what phenomenal states are like.

Recall that Nagel's purpose in introducing the "what it is like" lingo is to emphasize the subjective, experiential quality of conscious experience, and to show that we currently have no way of even understanding how such experience can be explained objectively. To quote Nagel:

> Very little work has been done on the basic question … whether any sense can be made of experiences having an objective character at all. Does it make sense, in other words, to ask what my experiences are *really* like, as opposed to how they appear to me? We cannot genuinely understand the hypothesis that their nature is captured in a physical description unless we understand the more fundamental idea that they *have* an objective nature (or that objective processes can have a subjective nature). (1974, p. 388).

If Nagel is right here, and I believe he is, then Kornblith must show how it is even conceivable to explain first-person experience in a third-person manner before positing that we engage in "theory construction" to understand what it is like to be someone. So, the theoretical link that Kornblith needs to make "knowing what it is like" possible in a third-person manner remains to be seen. Certainly, we can know things about a person through objective, third-person efforts, but unless Nagel's question is answered no such efforts can possibly tell us what it is *like* to *be* someone, as a conscious, experiencing agent, unless we happen to *be* the person in question.

This is not to say that Kornblith's skepticism about self-knowledge is entirely misguided, however. In his portrayal of skepticism, Kornblith mostly seems to have in mind general personality and character traits, such as friendliness, defensiveness, optimism, and so on. There is significant empirical work indicating that people can be very wrong about these things (Nisbett and Wilson 1977; Taylor and Brown 1988; Wilson 2002), and I think Kornblith is right to draw on these findings in order to understand the epistemology of introspection. I will discuss some of these empirical findings later, when we get to the conceptually-mediated introspection of one's own beliefs, desires, personality traits, and so on. For now, however, the main point to grasp is that these abstract psychological concepts are qualitatively different than phenomenal, conscious experiences, and that these different kinds of mental states are known about through very different means. As I have explained above, knowing what a phenomenal state is like is something that is immediately and intrinsically grasped by the experiencing agent, in virtue of the fact that the agent itself consists of the state. However, an agent comes to understand her general personality and character traits over time, through complex belief-mediated conceptualizations of her experiences and actions. This will be explained in more detail later, but for now it is sufficient to note that general beliefs about oneself arise through the application of folk psychological concepts and theories. So, unlike phenomenal self-knowledge, knowledge of one's own general psychological traits, such as one's beliefs and personality traits, are mediated and therefore prone to error and misconception. To

the degree that people tend to make false conceptualizations of themselves, some skepticism about self-knowledge is warranted.

However, in regard to the domain of phenomenal knowledge of one's own conscious states, skepticism is entirely inapplicable. When an agent has a phenomenally conscious experience, such as an orgasm, the agent thereby knows what the experience is like, immediately and intrinsically, in virtue of *being* the state. There is simply no room for error here. In fact, given the non-propositional nature of this kind of knowledge, it does not even make sense to ask whether it is accurate or not. Phenomenal knowledge cannot be wrong for the simple reason that it contains no propositional content. Now, the agent can go wrong in identifying or characterizing the experience, such as when a person misidentifies a surprising tap on the back as a pain, but this takes us into other, conceptually mediated, domains of knowledge. The conscious experience itself is what it is, and is intrinsically known as such.

4.6 Conclusion

In this chapter I have presented and defended the idea that phenomenal conscious experiences constitute a unique kind of introspective knowledge. We have knowledge of our own conscious experiences, in the sense of knowing what they are like, because we ourselves are composed of the experiences. We know our own conscious states by *being* them. This is the most immediate and primitive kind of

self-knowledge possible. In characterizing it, I have extolled its distinctive epistemic properties. The unique epistemic position obtained by being a conscious, experiencing agent needs to be recognized if we are to obtain a complete and accurate understanding of introspection, and I hope I have helped to bring about this recognition here.

However, it is also important to note the limits of this kind of self-knowledge. Phenomenal states in themselves have a very narrow and specific epistemic domain, consisting only of their intrinsic experiential qualities. They do not reveal anything about their metaphysical nature or causal structure, nor their conceptual relations to other states, either in the mind or out in the world. All of these things require further cognitive processes, whereby we can abstract from our experiences, conceptualize them, and construct propositional content about them. Identifying a phenomenal state as painful or pleasurable, reflecting upon an emotional experience and what it reveals about one's character, thinking about one's beliefs, and considering the course of one's life are all prototypical instances of introspection, and they all go beyond the immediate experience of phenomenal states in one way or another. So, phenomenal states constitute only a small portion of the broad range of introspective states. In the following chapters I will survey these further introspective domains, illustrating a variety of processes that extend beyond the intrinsic knowledge of experience to develop the more robust and self-reflective kinds of knowledge that are typically involved in introspection.

CHAPTER 5 – INTROSPECTION THROUGH FUNDAMENTAL COGNITIVE PROCESSES

5.1 Introduction

In the previous chapter, I explained an immediate kind of introspective knowledge that occurs among conscious beings through the intrinsically self-specifying nature of a conscious experience. However, many (if not most) introspective states are mediated through various higher-level cognitive processes. The typical understanding of introspection involves an inward-focused, soul-searching person who ruminates, analyzes, and reflects upon his or her own mental life. This takes us beyond the immediate knowledge of a conscious phenomenal state towards more complicated reflective states that engage the extensive cognitive capacities of the human mind. In this chapter I lay out a framework for understanding this more robust sense of introspection, by drawing upon the fundamental processes involved in human cognition. The central idea is that we engage in higher-level introspection by utilizing the mind's representational and information-processing capacities to represent and think about our own mental lives. Once again, there is no need to posit any special processes or mental abilities to account for introspection. The cognitive processes that we already know about (at least to some extent) can be utilized to understand what goes on when we introspect.

The full spectrum of cognitive abilities is relevant here: representation, memory, attention, learning, decision-making, problem-solving, conceptualization, language, and so on. However, to fully survey all of the cognitive capacities available to humans and analyze their roles in introspection would be a gargantuan task, taking us well beyond the immediate purposes of this dissertation. So, instead of providing a detailed explanatory account of all the cognitive processes relevant to introspection, my goal here is to offer a big-picture framework that outlines how we can understand introspection through the general cognitive abilities of the human mind. In doing so, I will focus on three basic cognitive capacities that I think are particularly important in accounting for introspection: representation, conceptualization, and attention. These fundamental, and overlapping, capacities of the human mind provide the basis for our capacity to understand the world we live in, thereby allowing us to understand ourselves as well. Taking a look at these capacities in broad outline will help us develop a satisfactory account of higher-level introspection, and will also lay a foundation for subsequent work in understanding the details of the various cognitive processes involved in introspection.

5.2 Representation, Self-Representation, and Metarepresentation

One of the most fundamental aspects of a mind is the capacity to represent. In fact, according to many philosophers of mind, the mind is in essence a representational device. For example, Fred Dretske proposes, as part of his

"Representational Thesis" about the mind, that "All mental facts are representational facts." (1995, p. xiii). In laying out the foundation of this representational thesis, Dretske offers a very clear and useful formulation of what constitutes representation:

> The fundamental idea is that a system, S, represents a property, F, if and only if S has the function of indicating (providing information about) the F of a certain domain of objects. The way S performs its function (when it performs it) is by occupying different states s_1, s_2,….s_n corresponding to the different determinate values f_1, f_2, ….f_n, of F. (1995, p. 2)

So, for example, a thermometer is a representational system in that it has the function of indicating the temperature of certain objects within a specified range. The states of the thermometer provide information *about* certain states of objects. Applying this to minds, the basic idea is that mental processes serve the function of indicating the states of various objects or events. Vision, for instance, has the function of indicating certain properties (movement, size, shape, color etc.) in an organism's external environment. Visual states provide information about various states of affairs, and are thereby representational. I take this as all being fairly standard and uncontroversial. Whatever else we may say about mental processes, they serve the purpose of representing things.

Dretske's approach to representation, and his application of this approach to introspection, is both important and useful, so I will be drawing upon it throughout this chapter. First, however, I want to make an important caveat in regard to a disagreement I have with Dretske's overall account. Doing so will help delineate to what extent his representational approach is applicable to self-knowledge, as well as clarify how the territory I will cover in this chapter relates to the issues discussed in

the previous one. According to Dretske, the representational understanding of mentality extends to the kinds of states we looked at in the previous chapter: phenomenal, conscious experiences. From this perspective, the way to understand pains, orgasms, color experiences, and so on, is in regard to their being informational indicators about things, events, or states of affairs. Dretske claims that this representational approach to phenomenal experiences explains their phenomenal character. For instance, in discussing the simplified, hypothetical example of a "mono-representational parasite", which embodies the sole representational function of representing the specific temperature of 18º C (indicative of a host for the parasite), Dretske states that:

> If you know what it is to be 18º C, you know how the host "feels" to the parasite. You know what the parasite's experience is like as it "senses" the host. If knowing what it is like to be such a parasite is knowing how things seem to it, how it represents the objects it perceives, you do not have to be a parasite to know what it is like to be one. All you have to know is what temperature it is. If you know enough to know what it is to be at a temperature of 18º C, you know all there is to know about the quality of the parasite's experience. (1995, p. 83)

So, according to Dretske, the representational facts about a mental state provide us with a complete understanding of what it is like to be in that state. This claim is misguided. First of all, it is at least conceivable (and probably true) that there are representations that have no qualitative experience at all. Let's stick with the example of temperature and consider a thermometer. A thermometer represents temperatures, but it seems unlikely that there is anything it is qualitatively like for the thermometer to be in one particular state or another. Of course, we can imagine that

lower temperatures feel cold while higher temperatures feel hot to the thermometer, but this is nothing more than a hypothetical projection of our own anthropocentric experiences. It does not rule out the possibility that there is simply nothing that it is like to be a thermometer, or Dretske's parasite. Secondly, it is also conceivable that the same objective feature may seem qualitatively different to different representational systems. For instance, we can imagine that Dretske's parasite may experience the temperature of its host in a number of different ways. Perhaps it feels like a pleasant tickle, or perhaps everything outside 18º C feels like a headache while the 18º C temperature of the host is simply the absence of the headache-like feeling. Through the consideration of such possibilities, we can see that the qualitative experiences of a representational system cannot simply be identified with the objective features of objects or events that the representational system represents. Even though we may know what is represented, the qualitative experience of the representation is still undetermined. So, Dretske's claim that if we know what a representational system represents, then we know everything there is to know about the qualitative experiences of the system, including what they are like (in the Nagelian sense), is false. Representational facts do not tell us everything there is to know about mentality.

This is not to say that qualitative experiences cannot be understood as representations, nor that what representations represent has no bearing upon what they are like experientially. Pains, color experiences, auditory sensations, and so on are representational states, and can be plausibly understood within a representational

framework. But, that understanding alone does not tell us what it is like experientially to *be* an organism that is composed of such representational states. There is a fundamental difference between understanding the functional nature of a representation and actually experiencing a state that can be understood in representational terms. This distinction is crucial here, and needs to be acknowledged. Otherwise, it can be erroneously conflated, leading to Dretske's incorrect and dismissive portrayal of phenomenal experience solely in representational terms. In fact, this is why I am treating these two issues in two separate chapters. The qualitative experience of being in a representational state and the functional / representational nature of that state may be the same thing metaphysically (arguably, a token brain state, at least in the case of us humans), but these are importantly different aspects of that thing, with very different epistemic qualities. Qua experience, the state is known in the existential sense I explained in the previous chapter, while qua representation the state is known in terms of the informational content it contains and the function(s) that information performs. Importantly, and as I emphasized in the previous chapter, the former does not fit into the standard epistemic analyses of whether or not it is veridical and / or reliable, as the kind of knowledge it engenders is not propositional, or informational, in nature. However, qua representation, mental states do fall under the standard epistemic evaluations, of being true or false, reliable or defective, and so on. We will consider the various epistemic qualities of introspectively relevant representations as we go along.

Now that we have this clarification out of the way, we can return to the nature of representation and how it pertains to introspection. We humans have a broad array of representational abilities, from our senses and the various representational capacities they involve to the conceptual representation of abstract entities and relations, such as justice, knowledge, causality, minds, and the speed of light. Of course, not all of these representational abilities are introspective. In fact, the majority of our representations are other-directed, representing the variety of things we encounter in our environment. However, we do represent aspects of ourselves as well, and in several different ways.

First of all, we represent states of our own body. Pains, for instance, represent certain instances of bodily damage (excluding cases of misrepresentation, of course). Another mode of body representation is proprioception, our general sense of bodily location. There is some disagreement regarding the nature of proprioception, but it is clear that we internally represent the location and states of many of our body parts (Sacks 1970; Ramachandran and Blakeslee 1998; Damasio 1999; Churchland 2002, pp. 59-126). Just close your eyes and silently move your arms (or other favorite body part) around. If you are like the vast majority of normal human beings, you will have information about the location and position of your arms. In doing so, you are engaging in a very basic form of self representation. Obviously, the representation of our bodily states to ourselves is vitally important, underlying our capacities to move limbs, maintain balance, and so on. In fact, as illustrated in figure 5.1, this basic capacity to represent the body is often regarded as a localized function of the primary

cortex in the human brain (although, interestingly, there appears to be some degree of plasticity within this representational capacity. See Ramachandran and Blakeslee 1998 for more on this.).

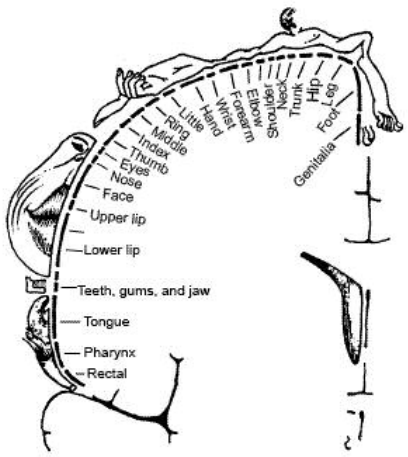

Figure 5.1 "Sensory Homunculus" from Penfield and Rasmussen (1950). This image illustrates the localized representation of body parts in the primary cortex, as discovered by Wilder Penfield.

So proprioception, along with other forms of representing one's own body, is a foundational aspect of our capacity to know ourselves (Bermúdez 1998). However, some may argue that representing one's own body is not a kind of introspection. Introspection, as it is typically understood, pertains to mental states. A representation of a bodily state, such as the position of one's own arm, is itself a mental state, but that which it represents is not. So, while bodily self-representation plays a fundamental role in self-knowledge, broadly construed, we need to move on to other kinds of representation, especially the representation of representations themselves, in

order to provide a satisfactory and comprehensive account of introspection, whereby we represent our own mental states.

It will be useful here to make some conceptual distinctions between the different ways in which we can represent ourselves. Representations, broadly speaking, can be about anything, their crucial feature being that they are *about* something, whatever that something may be. So, we can classify different kinds of representations according to differences in kinds of things that are represented. As discussed above, there are bodily representations, such as proprioception, pain, sensual pleasures, and so on. The defining feature of such representations is that they represent a physical state of one's own body. Alternatively, there are representations of mental states, such as beliefs and desires. (The distinction between mental and physical that I am drawing upon here is merely pragmatic, and is not meant to suggest that these are fundamentally different kinds of things. Arguably, mental states are just physical states of a certain kind). As discussed earlier, a central feature of mental states is that they are representational. So, considering the representation of these representational states leads us to the idea of metarepesentation, or the capacity to represent representations as representations (Dretske 1995, p. 43; Sterelny 1998; Sperber 2000). In regard to the broad range of representational states, metarepresentation can itself be broken down into different kinds: representations of beliefs, desires, sensory representations, memories, inferences, linguistic tokens, and so on. For now, however, I will consider metarepresentation more broadly, as the kind of representation that involves the representation of mental states. Before

turning to this, however, there is one further kind of representation worth mentioning, while we are drawing distinctions: the representation of oneself as a unified agent. In addition to representing various particular aspects and states of ourselves, we also represent ourselves as a whole, in terms of thinking, experiencing, and acting beings / selves. As in the case of metarepresentation, there are arguably a number of different subdivisions that fall under this category, such as representing oneself as a biological organism, a rational agent, or one of numerous kinds of social entities (parent, teacher, friend, voter, etc.). As we will see in the next section, these different modes of representing oneself as a self utilize different kinds of self-concepts, which may overlap in complex and interesting ways.

First, however, let's return to metarepresentation and the central role it plays in introspection. In essence, metarepresentation is the representation of a representation *as* a representation. So, for instance, if I think to myself that I am perceiving a computer screen right now, a component of that thought is the representation of my perception of the computer screen. In having this thought, I am representing my perceptual representation of the computer screen as a representation, thereby producing a metarepresentation. This capacity to represent representations is a fundamental aspect of introspection. Insofar as introspection involves the representation of one's own mental states (which, as we have already seen, can be understood as representations themselves), it inherently involves metarepresentation.

This more or less echoes Dretske's account of introspection. He states that:

> Introspective knowledge is knowledge of the mind—i.e., mental facts. Since mental facts (according to the thesis I am promoting) are

> representational facts, introspective knowledge is a (conceptual) representation of a representation—of the fact that something (else) is a representation or has a certain representational content. It is, in this sense, *metarepresentational*.
> … Introspective knowledge, being a form of representation, is, therefore, a metarepresentation—a representation of something (a thought, an experience) as a thought or an experience or (more specifically) a thought about this or an experience of that. If E is an experience (sensory representation) of blue, then introspective knowledge of this experience is a conceptual representation of it as an experience of blue (or of color). (1995, pp. 43 & 44)

So, when I introspect, say, that I have a pain in my left foot, I do so by representing the pain as a pain, through the mediation of a representational concept that characterizes the experience of pain as a pain. Along these lines, Dretske characterizes introspection as a species of "displaced perception," in regard to the fact that we do not know our own mental states by perceiving them, but rather by conceptually representing our perceptions (1995, pp. 41 ff.). For example, I know that I weigh 158 pounds not by directly perceiving my weight but rather by perceiving a scale and having a conceptual understanding of what scales represent, such that my perception of the scale is displaced into a "perception" of my weight via my conceptual understanding of the information presented by the scale. Analogously, I introspectively know that I see a computer screen not by perceiving my perception of the computer screen but rather by seeing the computer screen and having the conceptual capacity to characterize this perception as an instance of seeing. I represent my perceptual state to myself by displacing it into a conceptual framework that represents it as a representational perceptual state.

According to Dretske, such metarepresentational processes are simply all there is to introspection. As previously explained, I regard introspection as a heterogeneous collection of importantly different kinds of processes, so I do not follow Dretske in this reductive, single-faceted account of introspection. For instance, in regard to knowledge of pain, I not only know *that* I am in pain through the metarepresentational process explained above, I also experientially know *what* the pain is *like*, in the sense explained in Chapter 4. However, I emphatically agree with Dretske on the point that we do obtain introspective knowledge in this manner, and I think that this point is crucial to a complete understanding of introspection. We do have the capacity to represent our representations, and this engenders the ability to represent, and thereby know, aspects of ourselves. Moreover, this metarepresentational capacity, through the application of concepts and conceptual structures to ourselves, provides the foundation for more abstract domains of introspection, taking us beyond the relatively basic knowledge of our own immediate experiences (pains, visual perception, etc.) to the highly-conceptual knowledge of our character traits and social personas.

I will turn to the role that concepts play in introspection momentarily, but first there is an important epistemic point about representation that needs to be addressed. Wherever there is representation, there is also the possibility of misrepresentation. Representations, whatever they may be and however they may come about, are mediated through processes that can have errors. Experiences and information can be misconstrued or misinterpreted, concepts can be misunderstood or applied

incorrectly, and so on. Taking note of these possibilities, coupled with our understanding of introspection in terms of metarepresentation, we can see that introspective metarepresentation is fallible. Just as we can misrepresent a tree swaying in the wind at night as a person stumbling down the street, we can misrepresent a surprising tap on the back as a pain.

This runs counter to the common intuition that we cannot possibly be wrong about our own mental states, especially in regard to occurent states like pains and perceptual experiences. It may seem as though I cannot be wrong about being in pain, or about having a red visual experience, but insofar as introspection involves representation, this seeming is wrong. We can fail to accurately identify a sensation, misrepresenting it as a pain. We can incorrectly conceptualize a red experience as an orange one, and then continue on to believe that we saw something orange when in fact the color experienced was red. I grant that such mistakes may be rare and that the room for error may usually be small, but such misrepresentations do happen on occasion and that's all we need to dismiss the claim of infallibility. Representation makes misrepresentation possible, so it is possible for us to misrepresent our representations when engaging in metarepresentational introspection.

It is important that we do not overstate the case here, so allow me to more precisely state the epistemic impact of my denial of infallibility. Rather than being a complete mistake, the infallibility intuition is misguided only in regard to the particular domain of introspection currently under discussion. Alternatively, in regard to the immediate phenomenal knowledge involved in a conscious experience,

there really is no room for error. For instance, although I may mistake a pain for something else (such as thinking that a slight pain is a tickle), thereby misrepresenting my experiential state, that particular feeling of pain itself, as separable from any metarepresentational characterization of what the feeling is, constitutes an episode of phenomenal knowledge that cannot be anything other than what it is. Such experiential knowledge is what it is, irrespective of how or whether it is understood or misunderstood through higher-level cognitive processes, and so it is immune to representational errors. But, when we turn to represent our own experiences, the possibility of error immediately comes into play. Simply taking note of this difference between kinds of introspective knowledge can make considerable headway in clarifying the epistemic status of introspection, turning hopeless contention over whether or not introspection is infallible to more fruitful discussion of where and where not infallibility may legitimately apply. However, since metarepresentation is a foundational aspect of many varieties of introspection (basically, everything beyond immediate phenomenal knowledge), the fact that it opens the door for error is still very significant. Moreover, the more processing / conceptual mediation there is between the representation of oneself and whatever it is that is being represented, the more room for error there is. So, as we move on to higher and more abstract levels of introspection, the magnitude of representational fallibility will correspondingly increase. This will be important to keep in mind as we move on to the various modes of conceptually-mediated introspection.

5.3 Conceptualizing Our Own Mental States

Introspection through metarepresentation is enabled by having concepts of representational states. For instance, in order to represent the pain in my left foot as a pain, I must first of all have the concept of pain. Without this representational concept (or some other such concept), I would simply be unable to make the second-order step of constructing a representation of my representation *as* a representation. Once again, it will be instructive to look at what Dretske says about the matter:

> Not having a concept of representation does not prevent one from representing things, but it certainly prevents one from believing (hence, knowing) that one is doing it. Until a child understands representation, it cannot conceptually represent (hence, cannot believe; hence cannot know) that anything—including itself—is representing (and, therefore, possibly misrepresenting) something as F. What prevents small children and animals from introspecting is not the lack of a mysterious power that adults have to look inward at their own representations. They already have all the information they need, and they needn't look inward to get it. What they lack is the power to give conceptual embodiment to what they are getting information about. (1995, pp. 59-60)

Let's set aside the issue of whether or not children and animals actually have the conceptual abilities in question and treat this as a kind of thought experiment. Could an organism that lacks any conceptual understanding of representation be able to represent its representations as representations? Obviously, the answer is no. Representing something *as* X requires having a concept of X. For instance, consider the representation of an automobile as an automobile. In order to see an automobile and recognize it as such, one must not only have sensory experiences of the

automobile but also a concept of what an automobile is and the capacity to identify something that falls within that conceptual category. This is something that you and I do on a regular basis (presuming that you are a normal adult living in a cultural situation that is similar to mine), but someone or something else that lacks the concept of an automobile would be incapable of doing so (an animal, or an adult human that has been entirely isolated from the post-industrial world, for instance).

With this comparison in mind, let's return to the case of an organism that lacks any concept of representation. Such an organism simply would not be able to conceptualize its representations as representations, and so could not engage in the kind of introspection that we have been considering. This point in itself may seem to be so obvious that it is not worth stating, but it helps us identify what it is that enables introspection through metarepresentation: concepts. Concepts, whatever they may be, are the vehicles through which we organize our experiences and gather information into coherent, understandable bundles. Without concepts to structure them, our experiences would only amount to a "booming, buzzing confusion," to use William James' description of an infant's experience of reality (1890). This applies to our understanding of ourselves no less than it does to our understanding of the external world. Without concepts of representation, we can know the intrinsic phenomenal qualities of our experiences but we cannot form second-order representations of the experiences, such that they may become the objects of thought and understanding. So, considering the fundamental roles played by concepts in human cognition, a look at concepts will give us some insight into the nature of introspection.

Before jumping into the role of concepts in introspection, however, I want to address a potential worry that could creep up here. Pointing towards concepts to explain introspection may seem somewhat vacuous, especially to those who have explored the quagmire of views concerning the nature of concepts. There are a number of very different perspectives on what concepts are, from Jerry Fodor's claim that concepts are atomistic mental particulars to the more common, but still controversial, view that concepts are structured and interconnected representational complexes, not to mention the view that they are non-mental abstract objects (Fodor 1998; Margolis and Laurence 1999). So, without a clearer idea of what concepts are, appealing to concepts to explain introspection may not seem like a very progressive option. I don't have a definitive view of concepts to offer, so I will refrain from pressing any particular account of concepts here (with the exception of assuming that they are broadly mental in nature). Fortunately, however, we can still appeal to concepts without necessarily stating precisely what they are. Whatever they may be, concepts play central roles in human cognition and it is this functional level with which we are primarily concerned. Accordingly, we can discuss concepts in terms of what they do, rather than in terms of what they are. Of course, some views define what concepts are in terms of what they do, and this seems to me to be a fairly plausible route to take in understanding the nature of concepts, but for our purposes we can discuss concepts on this level without excluding the possibility that the nature of concepts includes more than their functional roles. Whatever they may be,

concepts enable us to engage in an amazing variety of cognitive processes, and it is this enabling capacity of concepts that I will address and appeal to in what follows.

With that said, let's get back to the meat of the matter. A central feature of concepts is that they allow us to engage in abstract thought, such that we can think about things beyond our immediate perceptual experience. Consider, for example, the concept of a cause. We utilize the notion of causation to understand a wide variety of things on a daily basis, from the satiation of hunger by eating food to the pressing of a gas peddle to make an automobile move forward. Yet, as Hume famously pointed out, we never observe causes, not even within ourselves. The concept of a cause does not come from, but rather extends beyond, our immediate experience, and thereby brings us to a new level of cognitive ability. It enables us to make sense of our experiences, by discerning abstract relations and patterns among them.

The very same sort of thing occurs when we employ concepts of mental representations to understand our own minds (and the minds of others). When we have the capacity to conceptually represent our mental states as mental states, such as when we conceptualize an auditory sensation as an instance of hearing or a feeling of hunger as a desire, we obtain a new level of cognition that enables us to understand our own mental lives. The abstract cognitive abilities enabled through concepts are especially important in this domain because, as I have already argued, mental states are not observable objects. Contrary to the perceptual model of introspection, we do not observe pains, perceptions, desires, beliefs, decisions, and so on. Instead, we

form representational concepts of them, such that they can be cognitively treated as objects. So, for instance, in believing that I believe I am currently sitting in front of a computer monitor, I utilize the concept of belief to abstractly represent my belief as an object that I can think about. Such metarepresentation through the employment of concepts is what enables us to engage in higher-level introspection, where we cognitively reflect upon our own mental lives.

Abstract, unobservable objects are typically regarded as theoretical entities, and this leads us to an important point about the role of concepts in introspection: The concepts of unobservable mental entities that we use in introspection are implicated in a broadly theoretical understanding of mentality. In other words, the utilization of one of these concepts involves some abstract, theoretical understanding of how the things that fall under the concept work. For instance, the concept of desire is typically connected to the abstract understanding that an entity with a desire for X will try to obtain X, all things considered. So, when I conceptualize a sensation as a desire, I do so with the understanding that the object of desire is something that I want to obtain. Similar explanatory frameworks surround our other concepts of mental phenomena, such as beliefs, pains, emotions, decisions, and so on. This theoretical understanding of mental phenomena forms a loosely-knit body of knowledge, known as "folk psychology" in the philosophy of mind and "theory of mind" in the cognitive sciences. Folk psychology / theory of mind and the role it plays in self-understanding (belief / desire explanations, in particular) will be the

focus of the next chapter, so for now I will only make a few general points regarding its conceptual / theoretical basis.

First of all, some clarification regarding the theoretical status of mental concepts is due here. In making this connection between concepts and theories, I am not necessarily suggesting that concepts themselves should be understood in terms of theories (for a review of this idea, and a comparison with other views on the nature of concepts, see the introductory essay in Margolis and Laurence 1999). Concepts may be understood in this manner, or they may be understood in some other way (or perhaps most plausibly, different kinds of concepts may require different kinds of explanations). As I said earlier, I am not making any claims here about what concepts are. Instead, what I am claiming is that concepts, whatever they may be in themselves, are intrinsically tied to theoretical understandings of the phenomena they encapsulate. Consider a concept of pain, for instance. Concepts of pain could be cognitively manifested in a number of different ways. But whatever it is, a concept of pain is clearly tied to some theoretical understanding of what pains themselves are and the roles they play in behavior, such that feelings of pain can be understood *as* instances of pain and individually distinguished from other kinds of feelings in virtue of that theoretical understanding.

Secondly, in characterizing our understanding of mental concepts in terms of theoretical structures, I am not necessarily suggesting that these theoretical structures are explicit, formalized theories, such as we see in contemporary science. Some have drawn parallels between the ordinary conceptual frameworks of humans and the

practices of scientists (see Gopnik and Meltzoff 1997, for instance), and I think that this is a worthwhile idea to explore, but I do not want to commit myself to a strong version of this claim here. Instead, in saying that our concepts of mental phenomena involve theories, I only mean to suggest that there are abstract explanatory structures that we utilize to understand, and perhaps also to predict, mental phenomena and their relations to behavior. This clearly has some overlap with the roles of theories in science. Scientists posit and use theories to explain and predict phenomena as well, and it is this general feature that is behind my use of the concept of a theory. However, scientists' particular methods in forming and using theories (developing explicit equations to codify a theory, for instance) are possibly quite different than what commonly goes on in a person's mind while introspecting (or while thinking about the mentality of others). So, when a person applies the concept of desire to her own feelings of hunger, and thereby understands her experience within a particular theoretical framework regarding what desires are and do, she isn't necessarily utilizing an explicit theory of desire, on a par with an explicitly formulated scientific theory, but she is using an abstract explanatory framework of some sort to understand that aspect of her own psychology.

Finally, although the labels "folk psychology" and "theory of mind" suggest a single, cohesive body of understanding, our theoretical understanding of mental phenomena may involve multiple, and perhaps even conflicting, conceptual frameworks. In regard to the prominent belief / desire explanatory framework that typically falls under these labels, it is plausible to think of folk psychology / theory of

mind as a somewhat unified theory. Again, this will be the topic of the following chapter, where I will further analyze the role this framework serves in self-understanding. But, in regard to the broader domain of mental concepts in general, our overall theoretical understanding is much more amorphous. In fact, there are different ways of conceptualizing at least some mental phenomena that can actually be at odds with one another. This is perhaps most apparent in regard to the self. The concept of a self is notoriously difficult to nail down to a single acceptable characterization, as those familiar with the literature on the topic will attest. Consider, for example, the conflicting intuitions at work in Bernard Williams' well-known article "The Self and the Future" (1970). Williams poses a variety of thought experiments which conjure up opposing views on the essential nature of one's self and how it persists through time. In some cases, it seems as though one's bodily continuity is most important to preserving the self, while in others one's self seems most closely associated with psychological characteristics, irrespective of bodily changes. There is much to say on this topic (see Kolak and Martin 1991 for a good anthology), but for our present purposes the point to grasp is that there are different conceptions of what a self is, and that these conceptions can come apart in certain contexts. We conceptualize the self in a variety of different ways: as a body, an experiential subject, a rational agent, and a social entity (which itself can be further broken down into a variety of conceptions: citizen, employee, spouse, and so on). These different concepts of the self often overlap, but they can diverge from one

another, in which case we can see how they implicate different theoretical frameworks for understanding the nature of the self.

In summation of this section, the primary point is that concepts of mental phenomena, along with their associated theoretical frameworks, provide the cognitive basis for higher-level introspection. The use of concepts to understand our own minds is not fundamentally different from our ordinary conceptualization of the world around us. There are concepts that play special roles in introspection (belief, desire, emotion, self, etc.), and there may be some self-oriented biases in the ways concepts are applied to oneself, but the same basic cognitive processes are at work in both one's conceptual understanding of the external world and oneself. The crux of this cognitive ability is the employment of abstract conceptual structures that allow us to make sense of our experiences. By applying concepts to our own mental phenomena, we bootstrap ourselves out of our immediate experiences into an abstract realm of concepts and theoretical structures, whereby we can understand and reflect upon what goes on in our own minds.

5.4 Attention and Introspection

No account of introspection could be complete without addressing the role of attention. Attention is intimately involved in the processing and organization of our perceptual data, actively directing our cognitive resources towards particular phenomena. Attention also guides our more abstract representational abilities,

involving the application of conceptual frameworks to our occurent thoughts and experiences. With these roles comes the capacity to direct our cognitive processes towards our own mind. For example, in order to become introspectively aware of a visual experience as an experience that you are having (suppose you are reflecting upon your experience of reading this text, for instance), you must attend to your experience and through that attention develop a conceptual representation of your experience as a visual experience of some sort. Without the capacity to attend to a specific phenomenon like this, no aspect of your experience could become an object of self-reflective understanding. So, clearly, attention plays a significant role in our capacity to introspect.

But, what is attention, really? As is the case with many psychological phenomena, there are multiple modes and levels of understanding what attention is. Consider the following diverse aspects of attention: 1) the first person experience of attending to things, 2) ordinary behavioral attributions of attention (for example, observing that one student is actively attending to a lecture, while another seems to be "somewhere else"), 3) the cognitive understanding of the functions involved in and performed by attentive processes, and 4) the neurological understanding of the actual mechanisms involved in attention. I will focus primarily on the first of these aspects of attention, but the others must be given their proper recognition. In fact, in some cases the third and fourth aspects may point out problems for the first two. This is an interesting and important issue, so allow me to digress for a moment, before addressing the role of attention in introspection.

We have an ordinary, common-sense understanding of attention that pervades our everyday discourse about the activity of our own minds and the minds of others. As William James put it, "Every one knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration of consciousness are its essence." (1890, p. 403). This seems intuitively compelling, as far as it goes. In our experience, our minds select some objects or thoughts to the exclusion of others, and through this selective process we have what seems to be a single, focused domain of conscious experience that we guide from one object to another. The metaphor of a spotlight is apt here, suggesting the idea that we direct our attention upon particular objects in the same way that shining a light upon objects highlights them, and makes them a focal point (see Baars 1997, pp. 95-111, for an application of this metaphor in understanding attention). But on closer analysis, it isn't exactly clear what is going on when we do this. For instance, there is evidence of multiple kinds of attention within the human mind, with distinctly different modes of operation (LaBerge 1995; Hardcastle 1998; Pashler 1998; Fuster 2002l; Itti et al 2005). This runs counter to our ordinary understanding of attention as a single, isolated focus upon some phenomenon. Arguably, there are even attention processes that we are not aware of, operating behind the scenes, so to speak, in the construction of our experiences and thoughts. Considering such things, our ordinary understanding of attention may not be adequate, or even on target, when it comes to understanding the actual cognitive processes involved in attention.

This concern should seem familiar by now, as it is similar to the more general point that I have been making about introspection, particularly in Chapter 2. Although we commonly understand introspection in perceptual terms (due to the utility of perceptual metaphors in getting a handle on our understanding of ourselves), this is not an adequate account of what really goes on when we introspect. Similarly, one may argue that our common-sense, metaphor-driven understanding of attention does not accurately map onto what is really going on in our minds.

Fortunately, however, there are important differences between attention and introspection (qua mental perception). As explained in Chapter 2, there is no identifiable perceptual mechanism whereby we literally perceive objects in our own minds, nor are there even identifiable mental entities to serve as the objects of such perception. Regarding attention, on the other hand, things are quite different. There are identifiable areas in the brain that can be associated with attention, and there is a variety of data that illustrates attentive processes in human cognition (ibid.). Moreover, since attention is a second-order cognitive process, there are no problems regarding the determination of what objects it takes as input (which, recall, was a devastating problem for the perceptual account of introspection). Rather than gathering information from determinable objects, as is the case with perception, attention is involved in processing already existing information in one's cognitive economy. So, considering the fact that these aspects of attention are much more readily identifiable than they are in the case of supposed mental perception, appeal to

attention as an explanatorily useful cognitive mechanism has at least some prima facie plausibility.

Unfortunately, I cannot pursue these issues further here. (It would take another dissertation to do so!) So suffice it to say that we can plausibly discuss attention as a legitimate category of human cognition, while still noting that we need to be careful that we don't take our common-sense understanding of it too literally. The actual processes involved in introspection are probably somewhat different than what they seem to be to our first-person conscious experience of attention, or to our ordinary attribution of attention to others, but this need not prohibit us from discussing attention without a detailed account of its nature, so long as we keep in mind the tenuous status of our understanding. Unlike the very dubious phenomenon of mental perception involved in a literal account of introspection, attention is a real phenomenon that we can both experience and empirically investigate.

With that said, let's return again to the role of attention in introspection, being mindful of the empirically revisable nature of what is discussed. It is worth noting that there are different aspects of attention that pertain to different kinds of introspection. To illustrate this, let's consider the concrete example of attending to a pain. First of all, in having a pain, one can attend to the feeling of pain itself. Arguably, this kind of attention plays a role in bringing about the phenomenal knowledge discussed in Chapter 4, but I am not going to enter into the difficult issue of how attention and consciousness relate to one another here. (For accounts that relate attention and consciousness, see Crick 1994 and Baars 1997.) Suffice it to say

that, whatever consciousness may be, paying attention to a state can play a role in making one phenomenally conscious of that state. A passage drawn from the perspective of Buddhist meditation illustrates the idea:

> If a sense object exercises a stimulus that is sufficiently strong, attention is roused in its basic form as a initial 'taking notice' of the object, as the first 'turning towards' it. Because of this, consciousness breaks through the dark stream of subconsciousness (-a function that, according to Abhiddhanna (Buddhist psychology), is performed innumerable times during each second of waking life). This function of germinal mindfulness, or initial attention, is still a rather primitive process, but it is of decisive importance, being the first emergence of consciousness from its unconscious subsoil. (Thera 1962, p. 24)

In addition to attending to an experience itself, one can pay attention to that which the experience represents, a scratched leg for example, in the case of a pain. With this mode of attention, the representational capacities discussed in section 4.2 come into play. Third, one can attend to the pain itself as a representation. As discussed in sections 4.2 and 4.3, this draws upon the metarepresentational capacity to conceptualize one's experience, by placing a particular experience into a conceptual category that explains or makes sense of the experience. Finally, one can also attend to oneself as the experiential agent that is having the pain, perhaps attending to the situation that one is in and the possible courses of action to be taken in response (removing one's leg from a thorny bush, for instance).

All of these different kinds of attention are relevant to various kinds of introspection. A great deal more could be said about this, mapping out the various connections between these mental processes, and perhaps my drawing attention to the issues will enable some progress to be made on the topic. For our purposes, however,

the interesting thing to note about these attention capacities is the general point that they can serve multiple cognitive purposes. Attention (whether it be the basic "bottom up" processes that filter and organize our basic conscious experience or the higher-level processes through which our abstract conceptual abilities are employed) can be directed towards either internal or external stimuli (or perhaps simultaneously to both), depending upon what aspects of the attended object are being attended to. In attending to a visual experience, for example, I can either attend to my experience of the visual field, reflecting upon the fact that I am having the experience, or I can turn my attention outward towards that which I am seeing, thinking about the various environmental factors involved in the production of the experience. Attention is an amazingly plastic capacity. Whether it involves multiple disparate mechanisms or the redirection of a particular mechanism (or some of both), the human capacity to attend to a multiplicity of phenomena exponentially expands the range of things that we can think about, including a variety of aspects of ourselves. Coupled with the ability to conceptually represent aspects of our mental lives, our ability to attend to various sorts of phenomena provides a foundation for introspective cognition. This is the primary point that I want to make about the role of attention in introspection. The very capacity to attend to various phenomena, however it is actually manifested in the mind, enables us to direct our cognitive resources upon ourselves. It is crucial to note that this introspective capacity originates from the general attention capacities of the human mind, and does not require appeal to any special capacity that is dedicated to

attending to ourselves. The ability to attend to ourselves simply falls out of the ability to attend to phenomena in general.

As with representation and conceptualization, it is important to note that the use of attention to engage in introspection opens the door for epistemic error. Attention helps us focus in upon aspects of our own minds, thereby making them potential objects of self-knowledge, but we can also fail to attend to aspects of our own minds (such as failing to identify the presence of a particular emotion) and perhaps even purposefully turn our attention away from aspects of ourselves that would otherwise be readily noticeable. So, the selective nature of attention is a double-edged sword, such that it both enables and inhibits our capacity to know our own minds.

5.5 Conclusion

In this chapter I have discussed how our ordinary cognitive processes are utilized in our ability to introspect. Arguably (from a broad evolutionary perspective), we have these cognitive abilities because they help us to take in and process information about our external environment. But, as we have seen, they can also help us understand aspects of ourselves. Perhaps, then, the introspective abilities that these basic cognitive processes engender can be understood as spandrels, or possibly exaptations, in that they are by-products and / or spin-offs of basic adaptive capacities created through the evolutionary processes involved in human ancestry.

Perhaps our introspective abilities are the by-products of our abilities to represent the mental states of others, where that ability has been vital for the kind of social cooperation that has played a crucial role in the survival and reproductive success of our species. Of course, this is highly speculative, and I don't want to put too much weight upon it, but it seems to me to be at least a plausible avenue that may be worth further exploration. In any case, the point to take home here is that, however it may have come about, the capacity of a human being to reflect upon its own mental states does not require the existence of any special cognitive machinery that is specially geared towards that task. To explain our introspective capacities, our basic cognitive machinery will do just fine. I have focused on the roles of three broad categories of cognition: representation, conceptualization, and attention, which are particularly relevant to our ability to introspect. But the real story of the cognitive processes at work in introspection is undoubtedly much more complicated than what has been covered in this general overview. I will move on in the remaining chapters to discuss some more specific aspects of the cognitive processes involved in introspection, as well as the epistemic implications they have. In addition, I hope that the framework I've provided in this chapter may help facilitate further inquiry into the nature of introspective cognition – inquiry that goes well beyond what I can cover in this dissertation.

CHAPTER 6 – UNDERSTANDING OUR OWN BELIEFS AND DESIRES: THE ROLE OF FOLK PSYCHOLOGY IN SELF-UNDERSTANDING

6.1 Introduction

We routinely utilize concepts of beliefs and desires when we identify, understand, and explain human thoughts and actions. This common understanding of beliefs and desires has come to be known as "folk psychology." Most often, the nature of folk psychology is discussed in terms of understanding others, from a third-person perspective. However, the domain of folk psychology is clearly applicable to first-person self-understanding as well, with significant aspects of introspection requiring appeal to and understanding of our own beliefs and desires. Considering this, the question emerges: What is the relationship between our folk psychological understanding of ourselves and of other people? In other words, how is it that we understand the thoughts and actions of both ourselves and others in terms of beliefs and desires? An intuitive first answer is that we develop an understanding of our own beliefs and desires through direct, first-person experience and then inferentially apply this understanding to others. In this chapter, I will bring this answer into question. I will defend the idea that our understanding of beliefs and desires, in regard to both others and ourselves, depends upon a general concept-mediated theory about human psychology. In other words, to have any understanding of our own beliefs and desires, we must (at least tacitly) rely upon a background theory that gives

meaning to the concepts of "belief" and "desire" as we apply them to the phenomena of our own mental lives. Although the view that folk psychology is a theory has been well supported (e.g. Churchland 1981 and Fodor 1987) and is now more-or-less the standard view in cognitive science (as indicated by common use of the phrase "theory of mind" to refer to the capacity to think in folk psychological terms), there is much less consensus regarding the issue of whether or not we use a theory of folk psychology in understanding our own thoughts and actions. I will elaborate upon this extension of the "theory theory" to understanding oneself, supporting it with empirical evidence and defending it against alternative viewpoints. Moreover, I will explore the epistemic implications that this view has for self-knowledge, illustrating how our knowledge of our own minds in this domain is significantly fallible.

6.2 The Common-Sense View

Consider the following dialogue between two university students:

Sally: Hey Sam, tell me something that you believe.

Sam: O.K. I believe that I have a head.

Sally: How do you know that you believe that?

Sam: Well, I've seen my head in mirrors hundreds of times. Plus, I can touch it.

Sally: No, no. You misunderstood my question. How do you know that your belief that you have a head is a belief that you have?

Sam: [with a puzzled look] Huh? It's *my* belief. How could I not know that it's a belief I have?

Sally: Let me try again. Why did you come to the food court today?

Sam: Obviously, I wanted something to eat.

Sally: So, you had a desire for some food, you believed that there is food available here in the food court, and so you proceeded to walk here.

Sam: Yeah.... [looking suspicious]

Sally: How do you know that's why you came here?

Sam: What do you mean, how do I know? *I* was the one that made the decision to come here, so obviously I know why. What more do you want? You philosophy people ask the most ridiculous questions!

I suspect that most people would sympathize with Sam here. For example, Kevin Flavey, seemingly speaking for the common-sense view expressed by Sam, states that "it is not uncommon, and frequently entirely appropriate to ask another how he knows what he claims about matters of fact in general. I assume that the singular inappropriateness of this question in the face of a person's avowal of his belief indicates that we have a distinctive warrant for accepting such avowals as true." (2000, p. 70). It seems intuitively obvious that we directly, transparently, and immediately know what we believe and desire. Moreover, although basing an act or choice upon what we believe and desire might involve a bit of inference, that too seems directly apparent to the acting or choosing agent. From this viewpoint, Sally's

questions about how a person knows his or her own beliefs, desires, and sources of action seem preposterous. A person has knowledge of her own beliefs and desires on the simple basis that they are *her* beliefs and desires. There is no need, or perhaps even possibility, of giving any further explanation or justification for one's knowledge of one's own beliefs, desires, and actions. They are, by their very nature, immediately given, and that's simply all there is to say about the matter.

It should be noted that on this view folk psychology can still be regarded as a theory. In fact, perhaps the most plausible formulation of this view is that it is because we have direct access to our own actions originating from our own beliefs and desires that we can come to a theoretical understanding of other people's actions on these terms. In other words, one might argue that our folk psychological theory is the product of inductive inference from our own first-person experience of our own actions to our third-person experience of the actions of others. Indeed, it often seems that this is what we do. For instance, I know that when I go to a restaurant it is usually because I have a desire for food and a belief that I can satisfy that desire at the restaurant. In observing someone else go to a restaurant, I inductively infer that the person probably has a desire for food and a belief that this desire can be satiated at the restaurant, on the basis of my past experience and its connections with similar behavior. At this point, I am making a theoretical inference, but one that seems rooted in my own first-person experience of analogous situations. From this perspective, my understanding of others is a matter of abstract conceptualization and theoretical reasoning, but my understanding of myself seems direct and unmediated.

As plausible as this perspective may sound, it is wrong. Sure, it seems intuitively obvious. But this is a case in which our intuitions lead us astray, giving us false confidence in the seeming transparency of our own minds.

6.3 The Theory Theory

The above view is wrong because our understanding of our own beliefs and desires is not direct and immediate, but rather mediated by a folk psychological theory. Instead of having direct access to our own beliefs and desires as sources of action, we tacitly apply theoretical constructs in attributing such states to both others and ourselves. This view of folk psychology as a kind of theory has come to be known as the "theory theory" (Credit has been given to Morton (1980) for first coining this term, but the initial idea goes back as far as Sellars (1956)). It should be noted that the term "theory theory" has been used as a label for a variety of specific views, some of which I do not necessarily want to be committed to here. For instance, Gopnik and Meltzoff use the label for their view that folk psychology (and other cognitive elements) is a product of processes akin to the construction of theories in science (1997). Although I think that this is a possibility worth exploring (as will be explained shortly), I wish to adopt a somewhat broader view of the theory theory in this essay. In essence, the view of folk psychology that I seek to defend here is that folk psychology inherently involves the use of abstract, theoretical concepts, however they may be acquired, in the attempt to understand (interpret or causally explain)

human actions. In other words, the attribution of beliefs and desires requires some theoretical understanding of what the concepts "belief" and "desire" entail. These concepts are not immediately given in our experience. Rather, belief and desire attributions involve the application of an abstract conceptual understanding of the phenomena they purport to describe.

This perspective is akin to the idea that observation is theory-laden, as discussed in the philosophy of science. Consider, for instance, Musgrave's depiction of this idea:

> ... when we look in the direction of the table and declare 'There is a table here' we interpret certain sensory inputs as indicating the presence of a table. But if we did not previously know the meaning of the word 'table', or possess the concept of a table, then we could not formulate the observation-statement. A creature which did not possess the concept of a table could still see the table; but such a creature could not learn from this experience that there is a table here and could not formulate the observation report 'Here is a table'. ... [This] gap between experiences and observation reports, a gap filled by language and concepts, is an important source of the fallibility of observation reports. (1993, p. 55)

In a similar fashion, I submit that our understanding of human thoughts and actions, including our own thoughts and actions, requires the employment of mediating concepts, namely, the concepts of "belief" and "desire", in conjunction with an abstract understanding of the roles beliefs and desires play in human thought and action. With this view of folk psychology, there is a crucial difference between having or experiencing folk psychological phenomena (i.e. beliefs and desires) and having a conceptual understanding of such phenomena. One can have the former without the latter. For example, although there is some room for skepticism here, I

think it is quite plausible to think of animals and children as having beliefs and desires. However, these creatures may have little (if any) understanding of their beliefs and desires *as* beliefs and desires (see Heyes 1998 for a critical review of this topic). This further understanding of beliefs and desires *as* beliefs and desires, which is pervasive in the world of human adults, requires a conceptual understanding of beliefs and desires that enables mediated inferences from the phenomena of beliefs and desires (e.g. the occurrence of having a belief or the experience of a desire) to a theoretical understanding of what these phenomena are and the roles they play in action. As Musgrave put it, there is a gap between an experience and the construction of an observation statement about that experience. The gap is produced by the jump from the mere presence of beliefs and desires to an understanding of such phenomena, and this gap is what makes our understanding of our own beliefs, desires, and actions a theory-mediated, opaque, and fallible phenomenon. So, contrary to the common-sense view, we do not have a direct and immediate understanding of our own beliefs and desires *as* beliefs and desires. Rather, we understand such phenomena via a mediated conceptual theory regarding the nature of such phenomena.

So what is this folk psychological theory, and where does it come from? In order to flesh out the terrain of the theory theory, I will briefly point to some possibilities here, without committing myself to any particular viewpoint. As already mentioned above, one view of the theory theory is that we acquire our theory of folk psychology in the same way that scientists construct theories. As expressed by

Gopnik and Meltzoff, the basic idea here is that "the processes of cognitive development in children are similar to, indeed perhaps even identical with, the processes of cognitive development in scientists." (1997, p. 3). From this perspective, we begin with an innate preliminary theory of human action and go through a variety of revising and reformulating processes in light of subsequent experiences, ultimately culminating in our adult folk psychological understanding. Another viewpoint, which has come to prominence via the development of evolutionary psychology, gives much more emphasis to the idea of having an innate theory of mind. According to this account, our folk psychology is rooted in an adaptive cognitive module that provides a "built-in" theory of human action (For one elaboration and defense of this idea, see Botterill and Carruthers (1999), pp. 77-103). In other words, it is a part of our "hard-wired" makeup to cognitively process human actions in terms of beliefs and desires. From yet another viewpoint, we acquire our folk psychological theory through our background society or culture. In illustration of this idea as it pertains to self-understanding, Lyons states that

> As soon as our "introspecting," or anything else for that matter, involves this web of intentional concepts [i.e. folk psychological theory], it is also inevitably involved in a certain stereotyped culture-based outlook. Thus our "introspection" of our motives, desires, deliberations, plans, decisions, or intentions will be colored by and adapted from our own particular culture's "rough-and-ready" view of the mind and its operations. (1986, p. 126)

As I see it, determining which of these viewpoints is correct is an open empirical issue. I am inclined to think that some modified combination of some or all of them might turn out to be the best account, but for the time being I am undecided on the

matter. Later, I will suggest some possibilities for deciding between these views through comparative social science research. However, the important point to note is what all of these views have in common. All hold that our folk psychological attributions, including our own beliefs and desires, are based upon some sort of theoretical conceptual structure. There is some sort of abstract cognitive processing (however acquired or instantiated) that mediates between our experiences and our understanding of those experiences in folk psychological terms. Once again, this is the core idea of the theory theory account of folk psychology that I wish to defend, as it relates to our self-understanding.

It is important to emphasize the significant contrast this view has to the common-sense view that people have of themselves. Instead of having a direct, immediate understanding of our own beliefs, desires, and actions, we actually apply a theoretical conceptual structure in understanding such aspects of ourselves. This means that our self-understanding is not necessarily special or privileged in this domain, but rather is often dependent upon the same general folk psychological understanding that we apply when interpreting the thoughts and actions of others. We know about our own beliefs and desires by applying the same conceptual framework that we utilize when attributing such states to others.

I should add, however, that this does not imply that we know about our own beliefs and desires in *exactly* the same manner in which we know about such states in others, through external, third-person observation. In virtue of being constituted by our own mental states (as explained in Chapter 4), we have first-person "access" to

our own mental states in a manner that is not available to others. For instance, I can attribute to myself a desire for water on the basis of a visceral, conscious experience of thirst, while I could attribute such a desire to someone else only by observing some behavioral indicator, such as seeing the person running to a water fountain or hearing the person say "I'm parched!" However, this is a difference in the evidential basis for attributing a mental state, and not a difference in the cognitive act of making an attribution. We can base our self-attributions on a variety of experiences that are simply not available to others, but the inferential act of attributing a belief or desire, on the basis of some "internal" first-person experience, requires appeal to the same concepts and folk psychological understanding that is used in third-person attributions. So, although there are notable asymmetries between first-person and third-person attributions of mental states, both kinds of attributions generally depend our theory of folk psychology and how well it matches up with actual psychological phenomena. Our self-knowledge concerning our own beliefs and desires, as well as their roles in determining the actions we make, is therefore significantly fallible, and contingent upon whatever theoretical structure we happen to be using. Now that I have outlined and described the basic position I hold, I will now turn to its defense, and will later return to its epistemic implications.

6.4 Empirical Support for the Theory Theory

As I see it, the most convincing case for the view that we tacitly apply a folk psychological theory when attributing beliefs, desires, and actions to ourselves can be made by appealing to empirical evidence. If we merely resort to a priori reflection, we are simply presented with the apparent immediacy of our beliefs and desires and are likely to remain within the paradigm of what I described as the common-sense view. In order to tease apart the seemingly transparent conception we have of ourselves and the third-person claim that we apply theoretical constructs when understanding ourselves, we must turn to some relatively recent empirical results. There are now a number of empirical studies that suggest that our self-attributions of folk psychological mental states are mediated, and can vary from our actual mental states. Consideration of these studies makes a strong case for the theory theory, so I will now discuss some of the most significant empirical findings in this domain.

Perhaps the most striking evidence comes from the now famous split-brain studies of severe epileptics. Gazzaniga, Sperry, and other researchers in the cognitive sciences have conducted a number of studies of people that have had their corpus callosum severed (as a remedy for severe epilepsy), with the two hemispheres of their brains thereby divided. One result of this procedure is that such people have two separate "sides" (for lack of a better word, and to avoid begging any questions about the nature of these people's minds) that are not necessarily aware of one another. This allows researchers to expose the two "sides" of a person to different stimuli,

without each side knowing what the other side experiences. The results of doing so are quite interesting, as Gazzaniga explains in the following scenario:

> One picture was shown exclusively to the left hemisphere and the other exclusively to the right. The patient was asked to choose from an array of pictures ones that were lateralized to the left and right sides of the brain. In one example, a picture of a chicken claw was flashed to the left hemisphere and a picture of a snow scene to the right hemisphere. Of the array of pictures placed in front of the subject, the obviously correct association was a chicken for the chicken claw and a shovel for the snow scene. One of the patients responded by choosing the shovel with his left hand and the chicken with his right. When asked why he chose these items, his left hemisphere replied, "Oh that's simple. The chicken claw goes with the chicken, and you need a shovel to clean out the chicken shed." In this case the left brain, observing the left hand's response, interpreted that response in a context consistent with its sphere of knowledge - one that does not include information about the snow scene. What is amazing here is that the left hemisphere is perfectly capable of saying something like, "Look, I have no idea why I picked the shovel - I had my brain split, don't you remember? You probably presented something to the half of my brain that can't talk; this happens to me all the time. You know I can't tell you why I picked the shovel. Quit asking me this stupid question." But it doesn't say this. The left brain weaves its story in order to convince itself and you that it is in full control. (Gazzaniga 1998, pp. 24-25)

For our purposes, the important point here is that the patient's left brain/mind does not have any access to the causal basis for his response, but still readily comes up with an explanation for the action. Presumably, the patient (the right side) chose the shovel because he believed that it most appropriately fit with the snow scene picture. The patient's left side was not aware of this and therefore could not appeal to this belief in explanation of his actions. However, as Gazzaniga emphasizes, this lack of access does not stop the patient from giving an explanation of his action. Without any hesitation, the patient confabulates a belief associating the shovel with the

chicken and falsely attributes this as the source of his action. How does he do this? The only plausible explanation, it seems to me, is that the patient tacitly made an inference based upon his general understanding of human psychology. In other words, he automatically drew upon a theory about what a rational person would have done in this case in order to explain his own action. Why else would a person pick a picture of a shovel when shown a picture of a chicken, besides appealing to some associational belief about cleaning chicken sheds? This seems intuitively obvious from a general folk understanding of human rationality (coupled with a cultural understanding of human use of chickens, of course), and it is such an understanding that this patient is applying to himself, just as he would to someone else. So, it appears that the human adult has an incredibly facile ability to reason about human actions via some sort of theoretical understanding regarding the causal basis of human actions, and that this ability is employed to understand not only other people's actions, but also one's own actions. This all happens quite automatically, such that it could easily give the appearance of direct access to beliefs and desires as the causal bases of one's own thoughts and actions.

One might worry here that this result is merely a strange by-product of people with split brains, and not an accurate depiction of what happens in the minds of normal people. After all, it seems that people with split brains might conceivably be thought of as two people, and therefore the above case may be more like a case of understanding someone else rather than oneself. However, this objection will not do

because the same sort of phenomenon can be found in normal, "unsplit" people as well. Consider, for instance, the following study conducted by Nisbett and Wilson:

> On a busy Saturday morning Nisbett and I placed a sign on a display table that read: "Consumer Evaluation Survey - Which Is the Best Quality?" We then made sure that four pairs of nylon panty hose were arranged neatly on the table and waited for the first passerby to stop and examine them. ... In an earlier version of the study, we noticed that people showed a marked preference for items on the right side of the display. We observed this same position effect in the panty-hose study. The panty hose were labeled A, B, C, and D, from left to right. Pair A was preferred by only 12 percent of the participants, pair B by 17 percent, pair C by 31 percent, and pair D by 40 percent, for a statistically significant position effect. We knew that this was a position effect and not that pair D had superior characteristics because in fact all the pairs of panty hose were identical - a fact that went unnoticed by almost all our participants. After people announced their choice, we asked them to explain why they had chosen the pair that they did. People typically pointed to an attribute of their preferred pair, such as its superior knit, sheerness, or elasticity. No one spontaneously mentioned that the position of the panty hose had anything to do with the preference. When we asked people directly whether they thought that the position of the panty hose had influenced their choice, all participants but one looked at us suspiciously and said of course not. (Wilson 2002, pp. 102-103; see also Nisbett and Wilson 1977)

I suggest that the subjects of this study, just as the split-brain patient described earlier, employed a tacit folk psychological theory in their understanding of their choices. Nisbett and Wilson offered a similar interpretation when they initially reported this study, stating that "when people are asked to report how a particular stimulus influenced a particular response, they do so not by consulting a memory of the mediating process, but by applying or generating causal theories about the effects of that type of stimulus on that type of response." (1977, p. 248). While they in fact had no good reason for choosing one pair of panty hose over another, the subjects of the

study readily attributed to themselves beliefs about desired features of the panty hose. It is highly unlikely that these attributed beliefs were genuinely present in the subjects at the time of their choice (unless they were already engaged in rationalizing their behavior at this point), as there were simply no distinguishable characteristics between the panty hose for such beliefs to be based upon. Instead, the best explanation seems to be that these beliefs about the panty hose were inferred by the subjects according to their folk psychological understanding of why people make such choices. In other words, the subjects explained their actions not by directly accessing the sources of their actions but rather by inferentially attributing certain beliefs and desires to themselves that make rational sense of such actions. Ideally speaking, rational people choose products on the basis of the products' qualitative features, as they relate to their desires and interests. The subjects in this study intuitively understood this, and applied this understanding to make sense of their own actions. I think that this provides significant empirical support for the theory theory, giving credence to the idea that we understand our own beliefs and desires via a tacit (but readily available) theoretical understanding of human psychology.

In case you are not yet convinced, studies in developmental psychology provide even further evidence for the theory theory. Through a number of studies, developmental psychologists have reached a general consensus that a fundamental shift regarding the understanding of mental phenomena occurs roughly around the age of 3 ½ in human children. For instance, 3 year olds often fail to have any understanding of false beliefs but by age 4 or 5 such an understanding appears to be

firmly in place (Gopnik 1993). For our purposes, the most significant feature of these findings is that 3 year olds generally make the same sorts of errors about themselves as they do about others. In an experiment pertaining to false beliefs, Gopnik and Astington had children encounter a candy box that had been deceptively filled with pencils. The experimenters questioned children after they had discovered the box's deceptive contents, asking them about both what other children will believe about the box and what they themselves believed about the box prior to discovering its contents. In both cases, the results were strikingly similar:

> One-half to two-thirds of the 3-year-olds said they had originally thought there were pencils in the box. They apparently failed to remember their immediately previous false beliefs. Moreover, children's ability to answer the false-belief question about their own belief was significantly correlated with their ability to answer the question about [another child's] belief ... (1993, p. 7)

It is important to note that memory effects were controlled for in this study. The children's inability to attribute prior false beliefs to themselves was not simply a matter of forgetfulness, but rather an inability to comprehend having a false belief. So, the children that failed in this task were missing some kind of understanding of human psychology, and they failed both in regard to others and themselves. Furthermore, the children that were successful at the task showed a significant correlation between attributing false beliefs to themselves and to others. These results strongly suggest that one's folk psychological understanding of oneself develops parallel to one's understanding of others. Rather than having some kind of prior or privileged access to our beliefs and desires, our understanding of ourselves is

on a par with our understanding of others, and is prone to the very same sorts of errors. Appealing to explanatory simplicity, our ability to understand our own mental states must therefore be dependent upon the very same cognitive processes we utilize in understanding the mental states of others. As Gopnik herself argues, this provides significant support for the theory theory:

> The developmental evidence suggests that children construct a coherent, abstract account of the mind which enables them to explain and predict psychological phenomena. ... Moreover, the child's theory of mind is equally applicable to the self and to others. (1993, p. 10)

So, we once again reach the conclusion that our grasp of our own psychology is mediated by some kind of theoretical understanding of human psychology. By considering the empirical findings I have outlined, we have found that the seemingly direct access we have to our beliefs and desires (as seen in the common-sense view) is somewhat illusory. Instead, our understanding of ourselves is theory-laden, being mediated through a background folk psychological theory. So, rather than having unquestioning confidence in our attributions of beliefs and desires to ourselves, understood in terms of direct and transparent access to the contents of our own minds, we ought to recognize that such attributions are only as good as the abstract conceptual devices we employ in making them. As I will again explain in a moment, this is not to say that we know about our own beliefs and desires in exactly the same way in which we know about the beliefs and desires of others, but rather that both kinds of attributions are rooted in, and therefore epistemically dependent upon, the same sort of abstract conceptualization. In understanding both our minds and the

minds of others, we apply a theoretical framework that makes sense of human experiences and behaviors (whether drawn from a first or third-person perspective) in terms of beliefs, desires, and the functions they perform.

6.5 Mind-Monitoring Mechanisms and the Phenomenology of Propositional Attitudes

Now that I have defended the theory theory as a central aspect of human self-understanding, I will consider some potential objections and, in doing so, also elaborate upon some important features of my perspective. Nichols and Stich (2003) have recently criticized the theory theory on the basis that it cannot account for self-awareness of one's own beliefs and desires. While they acknowledge that we employ a theory of mind in understanding the beliefs, desires, and actions of others, and even when we reason about our own beliefs, desires, and actions, Nichols and Stich argue against the idea that our initial access to our own beliefs and desires is a result of employing a theory of mind. Instead, they posit the existence of a "monitoring mechanism" that provides individuals with the ability to detect their own beliefs and desires:

> The basic facts are that when normal adults believe that p, they can quickly and accurately form the belief *I believe that p*; when normal adults desire that p, they can quickly and accurately form the belief *I desire that p*; and so on for the rest of the propositional attitudes. In order to implement this ability, no sophisticated Theory of Mind is required. All that is required is that there be a Monitoring Mechanism (MM) (or perhaps a set of mechanisms) that, when activated, takes the representation *p* in the Belief Box as input and produces the representation *I believe that p* as output. This mechanism would be trivial to implement. To produce representations of one's own beliefs,

> the Monitoring Mechanism merely has to copy representations from the Belief Box, embed the copies in a representation schema of the form: *I believe that* ___., and then place the new representations back in the Belief Box. The proposed mechanism would work in much the same way to produce representations of one's own desires, intentions, and imaginings. (2003, p. 13)

So, the basic idea is to posit some mechanism that transforms belief (and other mental) states into propositional representations of themselves. On this view, that is all that is needed for the detection of beliefs and desires. We do not need the application of a theory of mind to know what we believe and desire, but rather only a simple monitoring device.

I think that Nichols and Stich are correct to acknowledge that we do not necessarily access our own beliefs and desires in the same way that we access the beliefs and desires of others. As they point out, it would be absurd to think that we know of our own beliefs and desires solely by inferring them from our own externally observable behavior (2003, pp. 8-10). I take it as obvious that we access our beliefs and desires through some sort of internal mental processing that need not have anything to do with what might be inferred from observing our own behavior externally, out in the world so to speak. For instance, a person might identify her belief about the existence of some object (such as the presence of a bird in a tree) on the basis of certain sensory experiences, regardless of any externally observable behavior that might indicate such a belief. Similarly, a person could understand herself as having a desire for food on the basis of an experiential feeling of hunger, prior to and irrespective of any resulting behavior that may ensue from that hunger.

There are numerous first-person bases for mental state attributions that are not accessible from a third-person perspective, and any post-behaviorist account of human psychology ought to acknowledge this. However, I think that Nichols and Stich are too quick to jump from such uncontroversial claims about the data available to us from a first-person perspective to the idea that no theoretical structures are implicated in identifying our own beliefs and desires.

To see this, let us consider what Nichols and Stich's proposed monitoring mechanism really amounts to, which they say surprisingly little about. They propose that some sort of mechanical process takes a belief (or other mental state) and creates a further belief regarding the belief-holder having the belief. In order to do this, the proposed mechanism must, at a minimum, have access to some conception of what it is to have a belief, such that beliefs can be individuated from other mental phenomena. Otherwise, the proposed mechanism, however it might be constituted, simply could not identify a belief as a belief. Considering this, there must be some kind of tacit appeal to a conceptual theory of mind here. Although the concept of belief might be a relatively simple part of a theory of mind, it is nevertheless a theoretical posit, roughly on a par with other theoretical notions that play roles in our everyday understanding, such as the concept of an object or a force. To illustrate the plausibility of this point, consider again the significant difference between having a belief or desire and understanding that mental state *as* a belief or desire. Arguably, a dog might have the belief that sitting up on its hind legs will result in getting a treat without having any understanding at all that this state of belief is a belief. Such

understanding is something quite different from the presence of the belief itself. In order to have this further understanding, the dog would need to have some kind of concept-mediated grasp of what it means to have a belief, which is to say that it would need a theory that posits beliefs as a kind of mental entity, however this theory may be actually manifested in the mind / brain (perhaps by some quasi-mechanical process, as Nichols and Stich suggest). By taking note of this, we can see that Nichols and Stich's suggestion does not offer an alternative to the theory theory, but rather tacitly depends upon it for the identification of a belief as a belief. Whatever the merits of their proposed mechanism might be, it requires a tacit appeal to a theoretical conceptual structure that dictates what it means to have a belief, or other mental state. Without such a conceptual structure residing somewhere within a person's psychology (including the proposed monitoring mechanism), the idea of a person having a "belief box" is simply incoherent. There must be some kind of theoretical understanding in order for a person to form beliefs about beliefs in an understandable and utilizable manner.

Another potential objection to the theory theory comes from Alvin Goldman's phenomenological account of mental state concepts (1993). Goldman suggests that we identify mental states, including propositional attitudes like beliefs and desires, by phenomenological characteristics of types of mental states, and not by any functional role that such states play in a theory of mind. To this effect, he states that

> ... someone who had never experienced certain propositional attitudes, for example, doubt or disappointment, would learn new things on first undergoing these experiences. There is "something it is like" to have these attitudes, just as much as there is "something it is like" to see

> red. ... a plausible-looking hypothesis is that mental states are states having a phenomenology, or an intimate connection with phenomenological events. This points us again in the direction of identifying the attitudes in phenomenological terms. (1993, p. 24)

If Goldman is right here, then people can know about their own beliefs and desires without appealing to any folk psychological theory, by simply recognizing their phenomenological character. Notice that this view is a version of the common-sense view I described above. As Goldman puts it, "ascriptions to others, in my view, are "parasitic" on self-ascriptions…" (1993, p. 16). If correct, this perspective would count as a major mark against the theory theory as I have portrayed it, since it would exclude theory mediation in first-person access to folk psychological phenomena.

But is Goldman right? I think not, for a number of reasons. First of all, I find the idea of there being a phenomenological property that typifies beliefs to be highly dubious. People have beliefs with all sorts of different phenomenal qualities, from dread to delight and from interest to apathy. Moreover, many of our beliefs fall outside the domain of phenomenal experience altogether… tacit beliefs that do not make a direct appearance in conscious thought, for example. Considering both the variety of phenomenal qualities that may be associated with beliefs and the possibility of having a belief with no phenomenal quality at all, there is simply no identifiable qualitative experience that is specific to having a belief. Since Goldman's account requires that such a qualitative experience be manifestly apparent, I conclude that it fails. In contrast with Goldman, I am inclined to think that what typifies a belief as a belief is the role that it plays in one's psychology. When we call something a belief,

we do so in virtue of it being some kind of representative state that can form the basis for an inference or provide reasons for action. In other words, beliefs are beliefs in virtue of what they do in our cognitive economy. To adequately argue for this conception of beliefs would require much more than what I can plausibly do here, but, fortunately, my criticism here need not rest on this conception of beliefs. My present point is entirely negative, denying that there is any phenomenal quality by which beliefs can be identified. Whatever it is that makes a belief a belief, it is clearly not some qualitative state.

However, even if we grant to Goldman that there is a phenomenological character to beliefs, I still think that his account is fundamentally inadequate. There is a significant difference between simply having a phenomenological experience of a folk psychological state and identifying or understanding such a state in an explanatory, predictive, and / or interpretive manner. It is the latter that is important to our folk psychological understanding of ourselves. A phenomenological experience alone will hardly get you anywhere in terms of identifying your mental states, or understanding why you do what you do. To do these things, you must be able to accommodate the experience into some kind of theoretical structure that conceptualizes the mental state as the kind of mental state it is. To illustrate this with an example, let us consider the case of desire, which could much more plausibly be said to have a characteristic phenomenal state. Suppose that you are having a meal at an outdoor cafe. Just after your food is served, a rather skinny dog comes to your table and starts to whine. In this case, you would be more-or-less justified in

concluding that the dog has a desire for food. Doubts about animal consciousness aside, let's say that this dog has some basic qualitative experience typically associated with desire. But, could this dog understand its desire in the same way that you do? Could it identify its desire as such and understand the role the desire is playing in determining its actions? The answer could quite plausibly be no. In order to have this understanding of its desire, the dog would need to have access to some kind of conceptual structure that accounts for what desires are, at a minimum. This conceptual structure might include particular kinds of phenomenal experiences as identifying features of desires, but those phenomenal experiences in themselves do not convey that a desire is a desire. Such conveyance requires a concept under which the phenomenal experience may be placed. In other words, desires are not self-intimating; they do not in themselves present themselves *as* desires. Understanding a desire as a desire requires some additional cognitive machinery beyond the desire itself. So, having a phenomenal experience of a cognitive / folk psychological state does not necessarily entail that it is understood as such. As I have emphasized throughout this chapter, there is a significant difference between having or experiencing a folk psychological state and understanding the state in folk psychological terms. An understanding of a desire *as* a desire (or a belief *as* a belief) inherently requires not only the presence of a desire but also some kind of abstract conception of what a desire is. So, again, I conclude that Goldman's account simply falls short as an analysis of our folk psychological understanding of ourselves. The

theory theory must be brought into play in order for us to account for the manner by which we understand our own beliefs, desires, and actions.

6.6 Epistemic Implications

I will now turn to consider some of the epistemic implications of the view I have defended. If introspective reports of our own beliefs and desires are theory-mediated in the manner that I have described, then they may not be entirely dependable sources of information about our actual beliefs and desires, neither regarding ourselves nor others. As I have argued, our epistemic status in regard to our understanding of our own beliefs and desires is in general the same as our understanding of others. Both employ the same theoretical folk psychological understanding, and our thereby dependent upon the epistemic qualities of this theoretical understanding. Because our belief / desire attributions are mediated in this way, they are fallible. So, contrary to a long- standing train of thought in philosophy (largely following Descartes), we do not have direct, infallible access to the contents of our own minds, at least in regard to our understanding of our own folk psychological states. We can be wrong about our own beliefs and desires. Now, it is important to acknowledge here, as I did when discussing Nichols and Stich's account, that we do not necessarily come to understand our own beliefs and desires in *exactly* the same way that we do other people. We generally infer the beliefs and desires of others through their externally observable behavior, including their verbal reports, but

when it comes to ourselves we need not look at our external actions to discern such psychological states. We can use, and probably most often do use, our internal mental phenomena, such as our conscious emotional states, sensory experiences, memories, occurent thoughts, and so on. As I argued in Chapter 4, we have a unique kind of knowledge of our own conscious states, as they are known experientially. Because such information is only accessible to the individual having the states in question, this gives some limited credence to the common idea that we know more about ourselves than others. However, in regard to folk psychological phenomena, this is only a difference in the amount and kind of data used, and not a significant difference in the way that the data is used to generate folk psychological understanding. So, the basic point of this chapter still stands: We use the same theoretical structures that we use to understand the beliefs and desires of others to understand our own beliefs and desires. Although we do have a distinct kind of knowledge of our own conscious experiences, we can still be wrong about ourselves through the same fallible mechanisms that can cause us to be wrong about others.

Considering this, research that aims at achieving an accurate understanding of why people do what they do should not rely very heavily on introspective reports. For instance, suppose that Nisbett and Wilson's study regarding panty hose preferences had actually been an attempt to get at the real causes of people's choices in this domain. Obviously, the reports they gathered would tell them very little about this. In fact, the reports would be downright deceiving in this respect, as none of the subjects reported what appears to be a significant causal factor, namely the position of

the items on display. So, such reports cannot be relied upon in the attempt to uncover the "hidden springs" of human action. I do not have any concrete information regarding to what extent such reports are relied upon, but I suspect that they are solicited as data quite frequently in the human sciences. Anecdotally, I have encountered a number of studies that seem to take people's reports of their own beliefs and desires at more or less face value, but if the view I have defended here is correct, then this practice may be largely unjustified if the goal is to understand the actual causal factors behind human actions. Unless there is some indication that subjects are reporting their mental states accurately (which in some cases they might very well be… for some important findings that head in this direction, see Ericsson and Simon 1984), research cannot justifiably rely upon such reports to discern actual causal factors in a person's psychology.

However, if the goal of research is not to uncover the actual causal processes behind human actions, then I think that verbal reports can be given a more straightforward, and perhaps even elevated, role. For instance, if a researcher is interested in understanding the nature of people's own folk psychological conceptions of themselves (irrespective of whether they are accurate or not), then I think that verbal reports will be a highly valuable source of information to such research. In fact, it is hard to imagine what else might be a source of information for empirical research into such matters. I think that there is a lot of room for empirical development in this area. As mentioned earlier, there are a variety of views about the nature of folk psychology. For instance, some regard it as a cultural phenomenon

while others think of it as an innate cognitive apparatus. Further research regarding people's folk psychological reports about themselves might be an important source of information in these matters. I have in mind here a cross-cultural comparison of the sorts of folk psychological reports that people make about themselves (Lillard (1998) is an example of research heading in this direction, although with a focus on third-person attributions). Through such research, we could discern to what extent folk psychology varies from culture to culture, or person to person, and to what extent it is universal among humans. Just as Chomskian research in linguistics, which takes speaker judgments as a source of data, led to an understanding of language in terms of universal grammar structures that have a significant degree of cultural variation (among all the different languages of the world), it might be possible to use people's folk psychological reports about themselves to further our understanding of the nature of folk psychology, including finding out to what extent our folk psychological understanding is innate, cultural, and / or individually determined. Of course, this is highly speculative (and coming from a philosopher, it may not get much beyond that!), but I think that it is an avenue of research that could prove to be highly illuminating. In any case, the point for present purposes is that while our folk psychological understanding is fallible in regard to both others and ourselves, it might still be a highly valuable source of information in certain domains of the cognitive and social sciences, where the focus is upon the phenomena of folk psychology itself.

6.7 Conclusion

In this chapter I have addressed the relationship between our folk psychological understanding of ourselves and our understanding of other people, arguing that they both rely upon tacit theoretical structures. One's folk psychological theory provides the conceptual underpinnings necessary for attributing beliefs, desires, and actions to both others and oneself. So, our understanding of our own such mental states is fallible in the same way that it is fallible regarding the understanding of others. If we know our own beliefs and desires better than others, it is only because we have more data about ourselves (including, in particular, data from one's own internal psychology), and not because we have some privileged, unmediated access to these kinds of mental states. On this view, external, third-person perspectives (such as the experimental data I discussed) can even override a first-person report regarding the content of one's own mind or the reasons why an action was performed. While it certainly seems that we have direct, unmediated access to our own beliefs, desires, and actions, the empirical facts indicate that this is illusory. Instead, we apply a tacit theoretical understanding when attributing such things to ourselves. With this, the question emerges: Can we trust first-person reports about beliefs and desires? In light of the view I advocate, I have suggested the following answer: If we are looking for a genuine causal understanding of human psychology and action, then perhaps verbal reports do not have the privileged authority that often seems to be assumed. Such reports are only as good as the theory

they rely upon, and as we all know our common / folk theories about things can be misleading. In some cases, first-person attributions may be reliable, but this is something that must be empirically validated, rather than simply assumed. However, if the goal is to understand people's own folk psychological understanding of themselves (irrespective of whether this reflects an accurate picture of the actual cognitive underpinning of the phenomena), then verbal reports can continue to serve as a valuable source of information in cognitive and social science research. In fact, I suggest that such research might be developed to answer fundamental questions regarding the nature of folk psychology itself.

CHAPTER 7 – THE INTERNAL MONOLOGUE

7.1 Introduction

We find within ourselves a seemingly endless stream of words. Thoughts come and go in the form of linguistic expressions, whether we are thinking of what to eat for dinner, hashing out the intricacies of a philosophical issue, or rehearsing an intimate discussion with another person. In short, our consciousness is constantly accompanied by inner speech. In this chapter, I will discuss and analyze the nature of this inner speech and the roles it plays in introspection. I will argue that a significant portion of our thoughts (namely, our conscious propositional thoughts) occur through inner speech, and, consequently, that this vehicle for conscious thought provides the basis for substantial aspects of our awareness and understanding of our own minds. After explaining and defending this viewpoint, I will then move on to discuss the epistemic status of inner speech. I will suggest that a broad epistemic spectrum pertains to the phenomenon of inner speech, from utterances that self-constitutively determine their own truth value to the tendency of some linguistic phenomena to misguide, and perhaps even deceive, ourselves about the contents of our own minds.

Before beginning, it is important to discuss the connections between this chapter and the previous two chapters. In Chapters 5 and 6, I explained how much of our introspective self-knowledge is mediated by the higher-level cognitive capacities of human beings. In particular, I have given attention to the role of conceptual

understanding in introspection, where we apply abstract conceptual structures, such as the belief and desire framework of folk psychology, to interpret / understand / make sense of what is happening in our own minds. As will become apparent in this chapter, I believe that language plays a crucial role in this capacity to conceptually understand our own minds, insofar as such understanding is explicitly manifested in conscious propositional thought (unless otherwise indicated, I will use the term "language" to refer to natural language, such as English or Spanish). In essence, my view is that language is a vehicle with which we engage in abstract conceptual thought, through which we consciously entertain propositional assertions. To be clear, this is not to say that language provides the basis for thought simpliciter, but rather that it is one device (of, perhaps, many others) with which we think, particularly when we are engaged in conscious propositional thought about something (i.e. thought *that* something is the case, such as the thought "I am thinking."). From this perspective, language is an important ingredient in our ability to engage in the kinds of introspective thought that involve conscious conceptual representation of, and abstract reflection upon, our own minds. In other words, the introspective functions of language to be discussed in this chapter facilitate the conscious implementation of the introspective processes discussed in the previous two chapters. However, there is a reason why I have not mentioned the role of language until now. Although I believe that language plays an important role in introspective thought, the processes that I have discussed thus far could conceivably occur without language. From a purely analytic perspective, Chapters 5 and 6 can stand alone, without

necessary dependence upon the further claims about introspective thought that I make in this chapter. After reading this chapter, I hope you are convinced that these areas are, as a matter of fact, connected, but for those who remain dubious of the view that language plays a constitutive role in thought, I have arranged my analysis so that what I have said thus far does not stand or fall upon the roles I attribute to inner speech here in this chapter. With this background context in mind, let us now turn to consider the role of language in introspective thought.

7.2 Knowing Our Own Thoughts through Language

On the face of it, language appears to be a communicatory tool. We have a thought and then we express it in words. Relatively recently, however, several authors have proposed that language is more than just an expressive tool. It is, to some degree, a tool with which we think. In particular, it has been claimed that language, in the form of inner speech, provides the foundation for *conscious* thought. Without our inner monologues, where thoughts occur to us in linguistic form, we would not have the conscious thoughts that we in fact have. In this section, I will consider some interesting and important ways in which this claim has been argued for, which in turn will provide a foundation for understanding the role of language in introspection. In particular, I will focus on the work of Daniel Dennett, Peter Carruthers, and Ray Jackendoff in their various claims concerning the central role of language in conscious thought. Through the consideration of these views, I will

argue that significant portions of conscious thought are indeed rooted in language, and discuss the relevance that this fact has to a comprehensive understanding of our ability to introspect. However, I think it is a mistake to conclude from these considerations that language itself is the basis of conscious thought and experience. There are conscious elements of our minds that are fundamentally non-linguistic, such as basic perceptual awareness and the phenomenal experiences of emotions. Correspondingly, there are introspective domains that do not require linguistic utterances through inner speech, such as the intrinsic self-knowledge of conscious phenomenal states discussed in Chapter 4. Despite this caveat, however, I believe that Dennett, Carruthers, and Jackendoff do have it partially right, and can substantially contribute to a comprehensive understanding of introspection. Language does play a significant, constitutive role in much of human conscious thought, including the ability to consciously think about one's own thoughts. In particular, I will explain and defend the claim that conscious propositional thoughts (i.e. thoughts *that* X is the case) occur through linguistic activity in the mind. In other words, when we consciously think *that* something is the case, we are literally thinking with the inner speech locutions that contain the propositional objects of such thoughts as their content. From this perspective, inner speech constitutes a significant aspect of our introspective awareness of our own minds. In virtue of being the vehicle for conscious propositional thought, inner speech enables us to be aware of our own propositional thoughts, and, moreover, to consciously think about those thoughts.

Let us begin looking at the cognitive role of language by considering the views of Peter Carruthers. In his book *Language, Thought, and Consciousness: An Essay in Philosophical Psychology*, Carruthers proposes that natural language is constitutively involved, as a matter of natural necessity, in conscious thought. He claims that such thought consists in the tokening and mental manipulation of sentences in natural language (Carruthers 1996, p. 38). This is a version of what Carruthers calls the "cognitive conception" of language, which attributes a central role to natural language in human thought processes. This is opposed to the "communicative conception" of language, which conceives of language as merely a device for the expression of thought (Ibid., pp. 1-3). So, on Carruther's view, language is much more than something that we use for communicative purposes. It is an essential component of human cognition.

It is important to emphasize that Carruthers' version of the cognitive conception of language is confined to conscious thought. He believes that thought in general is conceptually independent from language, and grants the existence of thought without language. In explanation of this, Carruthers points to the thought-based behavior of animals and pre-linguistic humans (Ibid, p. 16). Such creatures do exhibit the presence of thought without natural language, through various elementary reasoning skills. However, Carruthers believes that *conscious* thought requires natural language, which leads to the controversial view that animals and humans lacking language are not conscious creatures. Later on, I will present some concerns that can bring this general viewpoint on consciousness into question. First, however,

I want to focus on the claim that the conscious thought of humans with language is constituted by their use of language, which, as we will see, can be distinguished from the broader claim about consciousness in general.

Carruthers grounds this claim in the phenomenology of introspection. As mentioned in my opening paragraph, we find within ourselves the constant presence of inner speech. Our thoughts occur in linguistic form. Based upon this observation, Carruthers goes on to argue that such internal linguistic occurrences constitute the thoughts we have. The words in inner speech simply *are* conscious thoughts. In support of this, Carruthers notes that we typically do not find any thought that is separate from or prior to an item of inner speech in our introspective awareness (Ibid, p. 50). There are simply the linguistic utterances of inner speech themselves. From this basis, Carruthers goes on to conclude that the items of inner speech themselves constitute our conscious experience.

But is this right? Consider the following observations of Otto, an imagined critic used by Daniel Dennett as a foil for explaining and defending his own views (which, by the way, reject the intuitive picture expressed by Otto):

> When I speak, I mean what I say. My conscious life is private, but I can choose to divulge certain aspects of it to you. I can decide to tell you various things about my current or past experience. When I do this, I formulate sentences that I carefully tailor to the material I wish to report on. I can go back and forth between the experience and the candidate report, checking the words against the experience to make sure I have found *les mots justes*. Does this wine have a hint of *grapefruit* in its flavor, or does it seem to me more reminiscent of *berries*? Would it be more apt to say the higher tone sounded *louder*, or is it really just that it seems *clearer* or *better focused*? I attend to my particular conscious experience and arrive at a judgment about which words would do the most justice to its character. When I am

> satisfied that I have framed an accurate report, I express it. From my introspective report, you can come to know about some feature of my conscious experience. (Dennett 1991, p. 230).

This perspective seems intuitively plausible, and it runs directly counter to the view of conscious introspection proposed by Carruthers. Rather than our conscious experiences being constituted by language, they are here seen as something prior to, or more basic than, language. Language is merely a communicative device, to express one's conscious experiences to others. These are strong intuitions. Consider, for instance, your perception of this piece of paper. You notice a particular shade of white, a particular arrangement of letters on the page, a particular shape and size, and all of these seem fundamentally outside any linguistic expression. So, how could our conscious experience be constituted by inner speech alone?

As I see it, the correct response is to limit Carruthers' account to conscious *propositional* thought. In accordance with standard philosophical usage, I use the term "propositional thought" here to refer to those thoughts that make some observation or statement *about* something, such that a propositional claim about the object of thought is made, as opposed to thoughts that are simply *of* things. For instance, suppose that you are looking at an apple. The mere awareness that you have of the apple is not a propositional thought, and can occur entirely outside any linguistic usage. However, if you have a thought *that* the apple is ripe, this would be a propositional thought. It involves making a propositional statement *about* the apple, taking your thoughts beyond perceptual awareness of the apple to an assertion about it. Considering this distinction, we can limit the domain of Carruthers' claim. Rather

than saying that language is constitutive of conscious thought across the board, it seems much more plausible to say it is constitutive of conscious propositional thought. With this, one can grant that basic perceptual awareness, and perhaps other modes of consciousness, occur outside of language. Only our conscious propositional thoughts, through which we reflect upon our experiences to think *that* such and such is the case, essentially involve language. In other words, our conscious thoughts that occur in language, which are quite prominent elements of human thought overall, are constituted by their occurrence in language. Inner speech is *a* kind of conscious thought, among, perhaps, a variety of others. In a 1998 article, Carruthers does in fact defend the claim that language is essentially involved in conscious *propositional* thought (1998). Although he does not explicitly say so, I suspect that it is something like the above concerns that motivated Carruthers to weaken his view in this manner. Regardless, I find this to be a much more defensible viewpoint than the initial, broader claim that language is required for conscious thought in general.

So, let us turn to Carruthers' defense of this weakened version of his view. According to Carruthers, if we do have conscious propositional thought at all, it must be through natural language. The only alternative is to be an eliminativist about conscious propositional thought, claiming that it does not exist at all (1998, pp. 460-461). Carruthers bases his argument here on the idea that to be a conscious object requires direct, non-interpretive access. In other words, the point is simply that a conscious thought is itself immediately present to awareness. But, if our inner speech is merely expressive of propositional thought, then propositional thought is mediated

through an interpretative device, and, therefore, not itself a conscious object. So, if our inner speech is not constitutive of propositional thought, then propositional thought itself cannot be a conscious object. The implication is that propositional thought, by default, lies outside our conscious awareness if it is not itself constituted by the inner speech that we are directly aware of. So, if we want to be realists about conscious propositional thought, taking our introspective awareness of inner speech as an awareness of our own conscious thoughts, it must be conceded that such thought is itself constituted by inner speech.

One response to Carruthers' argument here is simply to claim that we do have non-linguistic propositional thought that is directly and non-inferentially accessible in the manner required to be considered conscious. In illustrating this response to his argument, Carruthers notes psychological studies in which subjects claim to be thinking *that* such and such is the case without any corresponding linguistic token of thought (Ibid, p. 471). For instance, it seems intuitively plausible that I could be consciously aware *that* there is a piece of paper in front of me without tokening "There's a piece of paper" in inner speech. If this line of reasoning is correct, then conscious propositional thoughts can and do occur without language.

In response to this, Carruthers claims that the apparent presence of non-linguistic propositional thought is illusory. They are not direct perceptions of thought, but rather self-interpretations based upon folk psychology (Ibid, pp. 471-474). In making this response, Carruthers is very much in alignment with the view I put forth in the previous chapter, where attributing a propositional thought to

ourselves is understood to be much like our attributions of such thoughts to others, in the sense that they are mediated and thereby fallible. As we saw in Chapter 6, there are multiple psychological studies in which subjects construct false interpretations of their own behavior and reasoning. For instance, Nisbett and Wilson's study showed that, when asked to make a preferential choice between identical items, people will offer explanations for their choice that are clearly arbitrary and constructed after the fact (1977). While it might seem to people in such cases that they are directly accessing their thoughts, they are actually rationally reconstructing / confabulating their thoughts, based upon folk psychological inferences. So, according to Carruthers, the claim that propositional thought can only be conscious through natural language still stands. The presence of non-linguistic propositional thoughts is something inferred, and not consciously observed. The only objects that are actually available to conscious propositional thoughts are the linguistic occurrences that constitute such thoughts.

For my part, my introspective awareness of my own thoughts seems to confirm that conscious propositional thought occurs in and through the language of inner speech. An example will help illustrate this. In consciously noting *that* there is a computer screen in front of me, I am aware of the linguistic token "That's a computer screen" when the thought actually occurs to me. This inner speech act seems to me to be intrinsic to my conscious awareness of the thought, and nothing else is "there" in the thought besides the inner speech utterance itself. To be clear, this is not at all to say that my awareness *of* the computer screen is constituted by the

linguistic occurrence. I can be consciously aware of its shape, contents, location, etc., as purely perceptual objects, without uttering anything to myself. But, to think of the computer screen propositionally, to consciously think *that* the object I perceive is a computer screen, there are linguistic items that are intrinsic to my conscious awareness of the thought. Such inner speech seems especially prominent when I consciously reason about the computer screen. When I note that there is a smudged fingerprint in the bottom left corner of the screen and infer that I must have touched the screen at some point, these very word resound in my immediate awareness, and nothing else is present in the thought (besides my perceptual experiences *of* the screen, which occur without any discernable propositional content). I simply cannot imagine such a line of reasoning, *as a conscious object* in my own awareness, without such linguistic occurrences. So, when I reflect upon my own conscious propositional thoughts, language, in the form of inner speech, seems ubiquitous in my conscious experience of these thoughts. When I reason, think, reflect, speculate, etc., whether it be in regard to the "external" world or the contents of my own mind, the cognitive activity seems essentially embodied in language. I think in and with words. Not all of my conscious experience is linguistic, however. My conscious perception of the world through sense experience, along with my phenomenal awareness of being in / constituted by such experience (as discussed in Chapter 4), seems fundamentally unrelated to language. However, when I conceptually reflect upon that experience, forming propositional thoughts about it, I simply cannot imagine doing so without language. With these considerations, I find it reasonable to conclude with Carruthers

that the words embodying our conscious propositional thoughts simply are the thoughts themselves, and, consequently, that our awareness of our own inner speech constitutes our awareness of our own thoughts. It is when we reflect and reason about the world through formulating propositions about it that language becomes a central component of conscious thought, and it is through the occurrence of such language in inner speech that we become consciously aware of what we are thinking.

In addition, conscious propositional thought through inner speech can enable conscious thoughts about thoughts. This point brings us to one of the most interesting (and for our purposes, one of the most important) aspects of the relationship between language and conscious propositional thought, where language serves a recursive function in human thought and provides a basis for introspective thought through reflective reasoning. Daniel Dennett offers an intriguing "Just So" story that can serve to illustrate this function of language:

> Consider a time in the history of early *Homo sapiens* when language – or perhaps we should call it proto-language – was just beginning to develop. … Now it sometimes happened, we may speculate, that when one of these hominids was stymied on a project, it would "ask for help," and in particular, it would "ask for information." … Then one fine day (in this rational reconstruction), one of these hominids "mistakenly" asked for help when there was no helpful audience within earshot – except itself! When it heard its own request, the stimulation provoked just the sort of other-helping utterance production that the request from another would have caused. And to the creature's delight, it found that it had just provoked itself into answering its own question. What I am trying to justify by this deliberately oversimplified thought experiment is the claim that the practice of asking oneself questions could arise as a natural side effect of asking questions of others, and its utility would be similar: it would be a behavior that could be recognized to enhance one's prospects by promoting better-informed action-guidance. (Dennett 1991, pp. 194-195)

Although this is admittedly a highly speculative account, it highlights what could very plausibly be a central functional role served by inner speech in human conscious thought. On this story, language emerged as a means of communication. But, by turning language "inward", subjects could communicate with themselves, eliciting their own thoughts into conscious deliberation. Language thus becomes a means of questioning and answering oneself, bringing one's own thoughts to the fore, as consciously accessible objects. In this process, which Dennett calls "autostimulation," the mind obtains the capacity to interact with and reflect upon itself, without appeal to any sort of dubious internal perceptual mechanism (Ibid, p. 196). Whatever the merits or faults of Dennett's developmental account may be, it clearly highlights a fundamental function that is performed by the inner speech that we find so ubiquitous in our conscious experience. Through the linguistic utterance of our own thoughts in inner speech, we become consciously aware of our own thoughts, and, in turn, we can consciously think about these thoughts. As such, inner speech provides the basis for self-reflective thought and reasoning and can be regarded as a central cognitive function behind our ability to introspect.

Ray Jackendoff has argued along similar lines, maintaining that language can serve the purpose of bringing thoughts, including thoughts about thoughts, to conscious awareness (Jackendoff 1996, pp. 17-27). Language gives conscious form to what Jackendoff calls the "valuation of percepts" (Ibid., p. 25-27). The valuation of a percept involves noting / labeling phenomenal characteristics of a perceptual experience, such as noting that a particular percept is novel or familiar. According to

Jackendoff, having a language that can give conscious form to these valuations, in the form of linguistic utterances, enables us to consciously attend to them. For instance, uttering to oneself "Now that's something I haven't seen before!" when experiencing a novel visual perception brings that evaluative thought to the fore of one's conscious experience, in the form of the linguistic utterance itself. Moreover, such valuations can be attended to in themselves, providing the basis for recursive thought. For instance, I could categorize my thought about the novelty of my visual experience as a thought, uttering something like "I'm thinking about how this is a new experience" to myself, thereby making the thought itself an object of conscious thought. Jackendoff suggests that recursive linguistic valuations like this are what enable us to reason about reasoning. Reasoning, through the consciously attentive valuation enabled by language, can itself become an object of reasoning. In this manner, then, it is through language that we can reflect upon our own thoughts, making thoughts about our own thoughts consciously available. Consideration of this discursive process, coupled with Dennett's illustration of the "autostimulation" function of inner speech, highlights how language can be seen as an important constitutive element in our capacity to know our own thoughts. By labeling / conceptualizing our own thoughts through the consciously graspable medium of linguistic utterances in inner speech, we can consciously engage in thoughts about our own thoughts.

There are some significant potential objections to consider, but before addressing these objections it will be helpful to summarize the general viewpoint I have laid out in this section. I have argued for the view that the linguistic utterances

that occur to us in inner speech are the vehicles with which we engage in conscious propositional thought. Thinking to ourselves through the medium of language enables us to be aware of our own thoughts as they occur, and this in itself constitutes a kind introspective awareness, in that our own thoughts are manifested in a consciously accessible form. As argued by Carruthers, this is the only possible explanation for our being consciously aware of our own propositional thoughts; since we are aware of our own propositional thoughts through inner speech, if the thoughts themselves are something other than the linguistic occurrences in inner speech, then we are not consciously aware of the thoughts themselves. So, presuming that we are consciously aware of the thoughts that occur to us in inner speech, inner speech is itself the constitutive medium with which we think, and, simultaneously, with which we know our own thoughts. Moreover, following Dennett and Jackendoff, I have suggested that inner speech can serve a recursive function, allowing us to not only be aware of our own propositional thoughts, but to also engage in conscious thought about those thoughts. While I agree with Carruthers, Dennett, and Jackendoff in granting these constitutive roles to language in the production of conscious thought, I think that the further inference that language provides the basis for consciousness itself is misguided, which is a claim that all three authors make. I believe there are aspects of conscious experience that are fundamentally non-linguistic, such as the perception of sensory objects. However, this basic perceptual awareness is somewhat limited, encompassing only our immediate phenomenal awareness. In order to have conscious propositional thoughts about such perceptions, and thereby to engage in

conscious reasoning about them, I think that we need language as a vehicle to do so. Within this narrower scope, I think that Carruthers, Dennett, and Jackendoff are on to something in pointing out the cognitive functions of language as they pertain to our ability to know our own minds. Language is indeed a central component of human conscious thought, and taking notice of the roles it plays provides substantial insight into how we know (and, as we will see later, fail to know) about our own minds. This is a very significant point in understanding the nature of introspection, particularly the higher-level sorts of introspection that human beings are capable of. While we arguably share perceptual consciousness with other creatures, we are (apparently, at least) unique in having a complex natural language. It is this capacity that enables us to consciously think and reason in a propositional manner, and, in turn, to cognitively grasp the contents and events of our own minds. From this perspective, it is no coincidence that we, so far as we know, are the only creatures that have both complex natural language and that engage in abstract propositional thought about both the external world and our own minds. Language is the identifiable bridge between having conscious experiences and being able to consciously think about them in propositional terms, and, in turn, being able to think about such thought processes themselves as they occur in our own minds.

7.3 Objections and Replies

There are numerous objections to consider in regard to the view that language serves a constitutive role in human thought. I will mention and respond to what I take to be the most significant of these objections, as they pertain to the particular claims that language enables conscious propositional thought, and that the linguistic utterances that occur with inner speech provide us with conscious awareness of our own propositional thoughts. By addressing these objections, I hope to make the plausibility of my view more apparent.

One common objection to the view that we think with language begins with the idea that the same thought can be expressed in different languages, or differently in the same language. For example, it seems that if a person utters to herself "Ich liebe dich," she is having the same thought as she would if she had uttered "I love you" (in the same context). Both linguistic items are an expression of the thought that she loves this particular person. However, if, as I have argued, conscious propositional thought is constituted by natural language, then it seems to have the mistaken consequence that different linguistic items constitute different thoughts. Tokenings of "I love you" and "Ich liebe dich" are, in essence, different thoughts. Some might take this to be a reductio ad absurdum of my view.

While I am perhaps committed to different linguistic tokens being different thoughts, I do not think that this is as problematic as the above objection suggests. In response, I think it is important to note that what makes linguistic items seem to

express the same thought is that they have the same or similar semantic content. Different thoughts can be semantically related or even synonymous, while still being individually different thoughts. In other words, the individual utterances of "I love you" and "Ich liebe dich" can each be regarded as a single token thought, while still recognizing that they both fall under the same utterance type, in regard to their shared semantic content as a thought about love for another. Taking note of this token / type distinction adequately accounts for the understandable, but mistaken, notion that different linguistic items can express the same thought, rendering the fact that different linguistic items constitute different thoughts unproblematic.

Moreover, it is worth reiterating the fact that there is likely more going on in one's conscious awareness than the conscious propositional thoughts that one has. In the case of a person loving someone, the person likely has some conscious emotion through which that love is experienced, and this can happen without any linguistic events taking place. This conscious emotion may even have a role in the semantic content of the person's thoughts, but it itself does not provide the propositional structure of a conscious propositional thought, such as "I love you." On my view, it is only when the person has the conscious propositional thought *that* she loves the other person, whereby she consciously entertains a proposition about her experience via a particular linguistic utterance, that language plays a constitutive role in the thought. In other words, to say that the conscious propositional thought "I love you" is cognitively manifested through the linguistic utterance of those very words is not to suggest that the conscious experience of that love, or the full semantic content of the

thought, is reducible to the linguistic utterance alone. I suspect that many who oppose cognitive conceptions of language do so because they reject the reduction of human experience and mental content to linguistic phenomena alone. By offering my somewhat more nuanced account, whereby only specific domains of thought are identified with language, such opposition can be set aside, so that the genuine cognitive contribution of language, especially as it pertains to our introspective abilities, can be acknowledged.

Another potential objection is concerned with the appeal to introspective, first-person experience as evidence of how we actually think. One might notice that the line of reasoning I have given above, especially Carruthers' argumentation and my elaboration upon it, rests to a large degree upon what appears in our own first-person conscious experience. But, it might be questioned, what basis do we have for trusting such first-person experience as evidence for what is really going on in our minds? Couldn't we just be wrong or misguided in the introspective assessment that we think through inner speech? It seems to be at least conceptually possible that the inner speech I have argued to be constitutive of conscious propositional thought is really just a surface phenomenon (epiphenomenal, if you will), a reflection of deeper, unconscious, and perhaps innate, thought processes in the human mind. As I understand him, Jerry Fodor holds something like this view (or at least holds a view that can be extended in this direction), maintaining that thought is semantically prior to natural language, via a more basic, innate language of thought (Fodor 2001). The key idea here is that thought does occur in a language, but not natural language.

Natural language is a tool that expresses or communicates thoughts, not something that is essential to them. If this is right, then the introspective claim that inner speech constitutes propositional thought is illusory. Our inner speech is not constitutive of propositional thought; it is merely an expression or reflection of it. To put this point in the context of Carruthers' anti-eliminativist argument for conscious propositional thought, it is an acceptance of the idea that we may not be conscious of our own propositional thoughts as they actually occur in our minds. Perhaps in our inner speech we are only conscious of a mediated production of our propositional thoughts, and not the thoughts themselves.

Unfortunately, this is a possibility that I cannot rule out with certainty. However, I also cannot rule out with certainty the possibility that I am really just a brain in a vat, connected to a virtual reality machine that merely simulates my body and my experiences. In terms of pure conceptual analysis, these are possibilities that we will just have to live with, as epistemologists in the fallibilist tradition have come to understand. However, beyond acceptance of fallibility, there are some further considerations that can illustrate the plausibility of my viewpoint and save it from rejection via the somewhat eliminativist stance described above. Let us grant (for the sake of argument, if for no other reason) that there are propositional thoughts that occur in our minds through an innate language of thought, outside of our conscious awareness. Even if this is the case, we can still plausibly regard the linguistic utterances that occur through inner speech as propositional thoughts as well. First of all, there is no reason to think that these two kinds of propositional thought processes

could not coexist, and perhaps even interact with one another across the full spectrum of human thought. The human mind is a complicated, multifaceted device, and can arguably contain multiple modes of thinking, both conscious and subconscious. Considering this, we need not draw a mutually exclusive dichotomy between these two kinds of thought. We could have both subconscious and conscious propositional thoughts occurring in our minds, through different mediums, via our massively complex and parallel-processing minds / brains. Secondly, and more importantly, it would be very implausible to think that the linguistic occurrences in inner speech are not themselves thoughts. Full acceptance of such an idea would undermine our general sense of conscious, rational agency, whereby we explicitly engage in conscious thought and deliberation. To see this, think back for a moment upon the various thoughts you have had while reading this text. Presumably, you have been consciously thinking about what you are reading, and these thoughts of yours have been occurring through inner speech, whereby you utter thoughts to yourself in linguistic form. Now, if the eliminativist viewpoint described in the objection above is correct, all of this has to be wrong. Perhaps you have been thinking, but those thoughts would have occurred in a medium that you are entirely unaware of, and that which consciously occurs to you as your thoughts is merely an inert by-product. In other words, despite your conscious efforts to attend to and think about what you are reading, if the view under question is correct then there are simply no conscious thoughts happening in your mind. All actual thought is going on behind the scenes, and what you "think" you are thinking, via the words you encounter in your

seemingly conscious thought, is really not thought at all (I hope this is starting to sound preposterous to you, dear conscious, thinking reader!). From this perspective, what are we to make of all of those words that have been appearing to you in your internal monologue? They are not thoughts, and they are not performing any other discernable function. So, why are they there at all? To keep the conscious "you" entertained while your subconscious mind deliberates? I see no plausible way to account for these events, without rejecting the eliminativist perspective and accepting what seem to be our thoughts as thoughts. While these concerns do not produce a deductively valid reductio ad absurdum (whereby some logical contradiction in the view in question is exposed), I think they do show that it would be rather absurd to suppose that the "thoughts" that occur to us through inner speech are illusory, while all the real cognitive work is being done behind the scenes, so to speak, in a subconscious language of thought. To be clear, I am not denying that some, or perhaps even most, of the cognitive mechanics behind human thought happens on a subconscious, or even subpersonal, level, nor am I suggesting that the appearance of thoughts in inner speech provides us with entirely transparent and infallible awareness of our own thought processes (as I will explain in the next section, there is plenty of room for error here). All I am saying is that the thoughts that consciously occur to us in inner speech are indeed thoughts, and that the conscious accessibility of this inner speech provides us with the ability to know what we are thinking, at least under certain appropriate circumstances.

But, even granting that inner speech is not illusory, some may still have a problem with the idea that our conscious propositional thoughts are themselves linguistic. For example, in his recent essay "You Don't Know How You Think: Introspection and Language of Thought" Edouard Machery accepts the phenomenon of inner speech as real while still maintaining that "The introspective fact of inner speech cannot be evidence that our conscious thoughts are linguistic" (2005, p. 473). Machery bases his view on a distinction between vehicles of thoughts and contents of thoughts, claiming that inner speech is an imagistic representation of the content of our thoughts, and is not necessarily the vehicle of our thoughts. He illustrates this point through an analogy with visual images:

> Nowadays, not many people are willing to infer any property of the vehicles of these visual images from the properties of their content. For example, it is widely recognized that the visual image of a red apple does not have to be red: true, the visual image represents the property *red*, but that fact does not support the claim that the vehicle of this image also has that property. … In short, the sentences that are uttered in inner speech are represented by conscious images. They are part of the content of these thoughts. Now, generally, the fact that the content of some thoughts possesses a property P does not license *per se* the attribution of P to their vehicles. Hence, the linguistic nature of the sentences uttered in inner speech cannot be attributed to the thoughts themselves. (Machery 2005, pp. 475-477)

So, the basic claim is that inner speech is like the phenomenal experience of a color, in that it is an auditory image, and just as the experience of the color red tells us nothing about how we actually perceive color, the experience of inner speech tells us nothing about how we actually think.

This is an intriguing and initially compelling argument, but it fails to dismiss the view that we are consciously aware of our own thoughts through the language

presented in inner speech. Now, it is true that inner speech involves auditory images in some manner. As Machery points out, there are neurological studies showing that inner speech is manifested in the same brain areas that process language in inter-personal communication, where we physically hear speech through auditory perception (2005, p. 476). So, it is somewhat reasonable to assume that inner speech occurs through the experience of auditory images (incidentally, this is yet another example of how the mind utilizes already existing cognitive processes to engage in introspection, without need for any special process dedicated to that particular task). However, there are important differences between the imagistic experience of language and the perceptual experience of color images, such that Machery's analogy between these two phenomena does not hold up. In the case of language, the content is encoded in the vehicle in a much different way than occurs in our ordinary sensory modalities. A simple example should sufficiently illustrate this: I can convey my propositional thoughts to you through language, but there is no vehicle with which I can convey to you my experience of red. Amazingly, we can grasp the propositional thoughts of other people through language, but there simply are no analogous cases regarding our imagistic sensations, such as our experiences of sound and color. The transference of thought through language could not be possible unless the content of a thought is encoded in a discernible vehicle, which of course consists of the particular linguistic utterances that we understand when grasping a propositional thought through language. So, the vehicle and the content are intimately and constitutively connected together in the case of language, unlike perhaps anything else in our

experience. For this reason, Machery's analogy between language and color perception breaks down, and the distinction between vehicle and content that he draws from general imagistic perception does not apply to language. In the case of language, we can justifiably identify particular linguistic occurrences as the vehicles that encode the contents of thoughts. Otherwise, we could not feasibly account for our ability to understand the content of a linguistic utterance, whether it be a token of inner speech or speech from another person. Considering this, I think my account survives Machery's objection. We are consciously aware of our propositional thoughts through inner speech, because the language of inner speech is a content-encoding vehicle. It is worth adding that this does not imply that we know the cognitive mechanics of how such thoughts happen (which, arguably, is a legitimate target of Machery's argument). There are many complicated processes at work in language comprehension, most of which fall under the radar of our conscious awareness, but the point that we know what we are thinking via our inner speech utterances still stands.

Another potential objection to the view that language is constitutive of conscious propositional thought stems from concerns about explaining the intentional character of language and thought. For instance, John Heil has argued that the various positions typically held by philosophers on the relationship between language and thought are explanatorily defunct. Whether language or thought is given priority, Heil argues, the intentionality of both is left rather mysterious and unexplained, the intentional character of one being merely shifted over to the other (1988). On these

grounds, someone like Heil might complain of my account that it really does not get us anywhere in understanding how it is that we think *about* things. It is the intentional character of language and thought, the fact that they are *about* things that needs explanation, so claims simply about the relationship between language and thought tell us nothing truly informative.

I must admit that I am somewhat sympathetic to Heil's complaint here. Attempts to explain the intentionality of language or thought in terms of the other are not much more than diversions away from the real issue of understanding the nature of intentionality itself. However, explaining the intentionality of thought and / or language is not the task that I have undertaken here. My purpose has been to show that language is constitutively involved in conscious propositional thought, and to illustrate how this facilitates certain kinds of introspection, rather than explaining what language and / or thought are in themselves, qua their intentional character. In response to Heil, and in illustration of my sympathies to his position, I want to show how I see my account as being complementary to Heil's proposed endeavor of explaining the intentionality of language and thought together in naturalistic terms. As I have mentioned, I see language and conscious propositional thought as essentially related elements of human cognition. Rather than one taking priority over the other, I think that they are, in essence, elements of one and the same thing. Conscious propositional thoughts just are the tokenings of linguistic items in inner speech. Considering this, the possibility of explaining the intentionality of language and thought together becomes a viable enterprise, which is exactly what Heil calls

for. So, while my account admittedly has nothing to offer in explaining the intentionality of language and thought, I think it does pave the way for such an explanation to be given. It allows us to group important kinds of language and thought together, so that an explanation of their intentionality can be given in terms that are external to both. Of course, this is a project that extends well beyond the concerns I have addressed here. Nevertheless, the accord that my position has with such possible avenues of explanation and understanding should not be overlooked. Ultimately, I think that these views could potentially coalesce into a powerful and broad-reaching understanding of language and mental phenomena. For the time being, however, the point to take home is that language and conscious propositional thought, however they may instantiate intentional content in the mind / brain, can be understood as intrinsically united aspects of our ability to consciously think about things, including our own minds. Identifying the constitutive link between language, in the form of inner speech, and conscious propositional thought provides us with an insightful vantage point from which we can understand how we know our own thoughts and think about those thoughts.

A variety of additional objections and concerns could be raised regarding the relationship between language and thought, but I simply cannot address them all here. However, I would like to highlight some general points about my account that might allay some additional concerns that I have not explicitly addressed. Again, it is important to note the fairly narrow domain with which I am concerned. I have not made any general claims about the nature of language and thought altogether, but

rather made the more specific claim that natural language provides the basis for conscious propositional thought. There are arguably many other domains of human thought to which my claims do not apply. Moreover, even within this narrowed context, the claim I make is somewhat weak, in comparison to what might be said about the matter. Essentially, all that I have claimed is that when we are engaged in conscious propositional thought, the linguistic utterances occurring in our inner speech constitute the thoughts we have. As such, this mode of thought enables us to be consciously aware of our own propositional thoughts, and when recursively applied, to consciously think about our own thoughts. This does not exclude the possibility of other kinds of thoughts feeding into the content of these conscious propositional thoughts (such as memories, imagistic representations, emotional experiences, intuitive judgments, and so on across the broad palette of human thought and experience), nor does it exclude the possibility of various additional cognitive processes underlying our ability to engage in such thought (such as the numerous unconscious information-processing functions that are arguably operating behind the scenes to produce our conscious faculties). Considering these things, my view should be acceptable to all but the most adamant opponents of granting a cognitive role to natural language. In any case, even if you do not concede to the view I have offered, you may still be able to draw upon the further claims I make about the nature of introspection, by replacing the functions I grant to language with whatever cognitive mechanisms you think actually serve in those capacities. With that said, let us turn to

consider the epistemic attributes of inner speech, as they pertain to our abilities to know (and fail to know) our own minds.

7.4 Kinds of Self-Knowledge Enabled by Inner Speech

Now that I have explained and defended the view that we are consciously aware of our own thoughts, and able to think about these thoughts, through the linguistic utterances of inner speech, I will move on to discuss the epistemic properties of this mode of introspection. In this section I will describe three distinct kinds of self-knowledge that can occur through inner speech, and then in the next section I will discuss to what extent inner speech can lead us to truths about ourselves and to what extent it can lead us astray. In explaining the following three types of self-knowledge, I will also point out how the knowledge that occurs through inner speech overlaps with the kinds of introspective self-knowledge described in previous chapters.

As I see it, the knowledge engendered by inner speech through its constitutive role in conscious propositional thought is multi-faceted, and can lead to three distinguishable types of self-knowledge. First, the experience itself of undergoing inner speech provides the kind of phenomenal, self-constitutive knowledge discussed in Chapter 4 of our own conscious thoughts. For instance, when I think to myself "It might be cold outside. I should take a jacket," the very occurrence of those words in inner speech provides me with qualitative, experiential knowledge of the thought they

constitute. As already explained in Chapter 4, this type of knowledge is unique to the first-person perspective, in virtue of the experiencing subject being (partially) composed of that which is known. In other words, I know my particular thought about taking a jacket, via a linguistic utterance in my inner speech, in a unique manner because I myself consciously undergo the thought. This is a rather odd sort of knowledge, since (as explained in Chapter 4) it is non-propositional in nature and yet that which is known is itself a propositional thought. This may at first seem to present us with a paradox, but it is crucial to note that this is only one type of self-knowledge made available by inner speech, which may simultaneously overlap with other types of self-knowledge that have propositional content. With the understanding that self-knowledge is multi-faceted, noting the additional kinds of self-knowledge described below that accommodate propositional content, the seemingly paradoxical nature of this kind of self-knowledge can be rendered unproblematic.

The second kind of self-knowledge provided by inner speech is produced through the second-order function of taking our own experiences and other mental phenomena as objects of thought, such that they can be known in propositional terms. It is here that the functions of inner speech integrate with the cognitive mechanisms discussed in Chapters 5 and 6, whereby we conceptually represent our own mental lives. I leave it as an open possibility that there could be other modes of conceptual thought besides thought based in language, but language significantly contributes to our ability to engage in conceptual representation and is the medium with which we

consciously conceptualize our own mental states in propositional terms. For example, suppose that while on a diet you attend a birthday party where your favorite kind of cake is being served. Upon seeing the cake, you experience a desire to eat a piece of the cake, but then think to yourself "I really want a piece of that cake, but that would go against my diet. I should really just restrain my desire and go in the other room." In having this thought, as an occurrence in inner speech, you utilize the terms "want" and "desire" to conceptualize your experience, thereby making it an object of conscious propositional thought (which could, in turn, play a role in you overcoming the desire and accordingly controlling your behavior). It is through linguistic utterances like this in our everyday inner speech that we consciously engage in thought about our own mental lives. We use language as a tool to conceptually categorize, understand, and think about what is happening in our own minds, thereby developing a second-order domain of self-knowledge in which we can consciously know about our own minds in a propositional manner (e.g. knowing *that* we are in mental state X).

The third type of self-knowledge enabled by language pertains to the recursive capacity to consciously think about our own conscious thoughts. Not only can we conceptualize our mental states, as described above, but we can also conceptualize those conceptualizations, such that propositional thoughts themselves become the object of propositional thought. For example, suppose that you are an undergraduate student who was raised in a religious community and thereby acquired a belief in God. You are now sitting in an introductory philosophy course and the

instructor asks "How many of you believe in God? Raise your hands." You think to yourself, "Well, I've never really questioned it, but I've always been taught that there is a God. So, I guess I do believe in God" and then subsequently raise your hand. At this point, you have consciously conceptualized your belief in God as a belief, and are thereby engaged in the second kind of self-knowledge described above. After a few moments of pacing back and forth in front of the class, your philosophy instructor then says "Most of you appear to believe in God, but why do you? Are there good reasons for doing so? To get things started, let's consider the problem of evil. Evil things happen on a near-constant basis in our world, from disease and natural disasters to the murder of innocent people, including helpless children. How could this be if there is a God, as an all-good, all-knowing, and all-powerful being?" The lecture goes on, but this is enough to get you reflecting upon your belief in God. You think to yourself "Well, why do I believe in God? I guess it's because that's what I've been taught in church and from my parents. That's what came to mind when he first asked about it. Is that a good reason? Well, maybe not since they could be wrong, like that Descartes guy said. And there's this problem of evil. When I think of God, I do think of him being all-good and all-powerful, but if he is then why is there evil in the world? If God exists wouldn't he put an end to evil? Maybe there's something wrong here with the thought that God is all-good and all powerful…" And on your thoughts could go, consciously reflecting upon your thoughts about what you believe, perhaps even augmenting your beliefs in light of the reasoning you engage in through this consciously self-reflective thought process. Through this example, we can see

how inner speech locutions are utilized to engage in conscious propositional thought about thought, such as the consideration of beliefs about one's beliefs. This brings us to the highest capacities of the human mind (highest in a descriptive sense only, without necessarily importing any normative considerations here), as they pertain to our ability to introspect, whereby we can know about ourselves through abstract representation of and conscious thought about our own thoughts. As I have suggested, inner speech is the cognitive basis for extending introspective knowledge into these multi-level and recursive domains of conscious thought. Through such language-enabled thought, we can consciously reflect upon the content, processes, and nature of our own minds, thereby developing a multi-faceted conceptual understanding of who and what we are that culminates in the first-person perspective commonly regarded as introspection.

7.5 From Self-Determined Truth to Self-Deception and Back: The Epistemology of Inner Speech

Now that we have identified the major types of self-knowledge enabled by inner speech, we can move on to consider some more specific epistemic qualities of inner speech, in terms of what propensities it has to produce true or false beliefs about ourselves. As I see it, the epistemic territory is quite diverse here, ranging from utterances that self-determine their own truth value, simply in virtue of their occurrence, to tendencies to engage in narrative rationalizations that can produce false

beliefs, and even self-deceptive assent to false beliefs, about oneself. Considering this broad range, a full treatment of the issues would require more than I can offer here, but given the nature of this project it is important to define the general territory and highlight some of its more prominent features. I will point readers to further resources and discuss the need for further research as is appropriate.

Let us begin with the kinds of utterances that determine their own truth value. Suppose that you are at a new restaurant for the first time, looking over the menu. After pondering over a few items you stumble across "Mama's Special". This is a dish that you have never encountered before, but the ingredients sound delicious to you. You think to yourself "Sounds good. I think I'll have Mama's Special." In this case, your particular thought constitutes self-knowledge of what you think you want to order at that particular moment. Moreover, the very occurrence of the thought is what makes it true. The thought "I think I'll have Mama's Special" is itself a conscious thought of a desire to have Mama's Special, and as such its occurrence makes itself true (of course, this is assuming that the concept of a "thought" is a viable psychological category to begin with). This is true even if you harbor a hidden desire to order Daddy's Chili instead and / or you end up just eating from the salad bar. At the moment that the thought occurs to you, when you are consciously thinking "I think I'll have Mama's Special," you are in fact thinking that you want Mama's Special, irrespective of what else might be going on in your mind. Considering this, we can see that some utterances provide self-constitutive propositional knowledge that does not match up with the standard epistemic

paradigms used to evaluate typical propositional knowledge (i.e. knowledge about the "external world"), in terms of evidential support, reliable perceptual mechanisms, and so on. When you utter to yourself "I think I'll have Mama's Special," you are not necessarily uttering a proposition that is true in virtue of mapping on to some state of affairs that independently exists, external to the utterance. Instead, the utterance of the proposition itself provides the content for determining its truth value. So, inner speech utterances of this type have a somewhat unique epistemic quality, in that they are the truth makers of their own propositional content.

This leads us into the murky but interesting territorial overlap between human agency and the epistemology of self-knowledge. In so far as we have some kind of genuine agency, through which we actively engage in the creation of our own mental states, knowing ourselves is less like evidential discovery and more like a self-constitutive, constructive process. Of course, in terms of epistemic theory, this is also like opening a big can of worms, an event which I do not want to get embroiled in here. This territory has recently been taken up by a handful of philosophers, so interested readers can turn to the following sources for more developed and nuanced discussions of this matter: Richard Moran's *Authority and Estrangement: An Essay on Self-Knowledge* (2001), David H. Finkelstein's *Expression and the Inner* (2003), and Dorit Bar-On's *Expression and Self-Knowledge* (2004). Let it suffice to say here that in a certain range of cases, the conscious propositional thoughts that we have about our own minds make themselves true, such that we know these aspects of ourselves simply in virtue of the fact that we ourselves are the causal agents that

define their content, through the utterances we make in inner speech. Although this may only constitute a small part of introspective self-knowledge overall, it needs to be included in any comprehensive account of what and how we can know about our own minds.

However, acknowledgement of the above sort of self-determining self-knowledge is not to say that we simply make up the content of our own minds, such that whatever we say about ourselves becomes true. There are other cases where we can utter things to ourselves that are not true of our own mental states. In fact, the majority of our propositional self-knowledge is arguably fallible, such that uttering a false propositional statement about oneself is at least possible in principle. As discussed in Chapters 5 and 6, the kind of self-knowledge where we engage in conceptual representation of our own mental states is mediated by processes that could conceivably go wrong, thereby making room for epistemic error. In other words, when we conceptually represent our own mental states as being a certain way, there is automatically a gap between that which is represented and the representation of that thing, such that the possibility of misrepresentation arises (with the exception of cases where the representation is itself the represented object, such as the "I think I'll have Mama's Special" example above). Considering that inner speech is a means of engaging in conscious propositional thought about our own mental states, through which we conceptually represent our mental states, inner speech is subject to this sort of epistemic error. Once again, an example will help illustrate my point. Suppose that you are one of the unwitting subjects in the Nisbett and Wilson study I discussed

in Chapter 6 (Wilson 2002; Nisbett and Wilson 1977). You are out at a shopping mall and come across (what appears to be) a panty hose customer survey booth. You try to simply walk past it, but the booth attendant insistently asks you to look over 4 pairs of panty hose laid out on a table and pick which pair you like the best. You casually inspect each pair, from left to right, and while doing so you note various characteristics, thinking such things as "that feels soft," "seems very elastic," and so on. You pause for a moment after inspecting all four pairs, but without much conscious deliberation (because, of course, the choice does not really matter much to you) you say "I like D the best," pointing to the pair on your far right. When the attendant asks you why you chose that pair, you think to yourself "well, I remember noticing how smooth it is" and then say out loud "because it's the smoothest". The attendant then asks you if the position of the panty hose had any impact on your choice. You think for a moment "Hmm the position… I only remember thinking about smoothness and elasticity" and then answer "No." You walk away and your conscious thoughts about choosing on the basis of smoothness quickly turn to other, more personally immanent matters. In this case, your conscious thoughts in inner speech represented your choice as being based upon qualitative differences between the pairs of panty hose. However, in this particular case the panty hose are all qualitatively the same! Moreover, as Nisbett and Wilson found, there was a significant position effect, such that a statistically significant number (40%) of the participants chose pair D on their far right (and another 30% chose pair C, in the mid-right position). So, it is quite plausible to think, from a somewhat objective third-

person perspective, that your choice was causally produced through some tendency to choose items on the right. Perhaps, for instance, you are right-handed and have a propensity to be drawn towards things to your right (despite a lack of conscious recognition of this propensity). The point here is not necessarily that this was in fact the causal basis of your choice, but rather only that it is conceivable that such is the case. It is quite possible that your conscious representation of your choice, through the inner speech you engaged in while going through the process of choosing a pair of panty hose, was mistaken and did not accurately reflect the mental states that actually produced your choice. In short, your conscious propositional thoughts about your choice, as embodied in your inner speech, are fallible.

Evidence of such fallibility can be inferred from outside the phenomenon of inner speech as well, among studies of the accuracy of verbal reports. Psychologists have identified a phenomenon called "verbal overshadowing," where the soliciting of verbal reports can significantly decrease memory accuracy (Schooler and Engstler-Schooler 1990; Meissner and Memon 2002). For instance, Schooler and Engstler-Schooler found that observers of a video robbery scene who were subsequently asked to write down detailed descriptions of the robber's face were much less likely to accurately identify the robber's face at a later point in time than where observers who did not engage in any overt verbalization of the event (with 38% accuracy among the verbalizers and 64% accuracy among the non-verbalizers), and similar results have been found in numerous other studies (Ibid.). So, the use of language can actually impede our ability to accurately access our own mental states, such as our memories.

Although the subjects in such studies engage in overt verbalization, it is reasonable to assume that similar effects can occur in our inner speech commentaries upon the contents and attributes of our own minds. This provides some fairly strong empirical support of the fallibility of inner speech in its role of producing introspective self-knowledge.

With the above considerations, it should be sufficiently clear that inner speech can lead us to make mistakes about the nature and content of our own minds. It should be added, however, that it is not yet clear to what extent the verbalization of our mental states and processes has a tendency to produce inaccurate results. Although there are fairly robust results demonstrating decreased accuracy in mental state attributions among verbalizing subjects, there is also some significant evidence suggesting that verbal reports can be reasonably relied upon in some contexts. Most notably, Herbert Simon and K. Anders Ericsson have developed protocol analysis methods designed to avoid the misguiding tendencies of verbalization (Ericsson and Simon 1984; Ericsson 2003). Setting aside some of the nuanced details, the main idea is to solicit verbal reports with very specific directions to verbally state one's thoughts as they occur in real time. Under such circumstances, Ericsson states that:

> subjects can think aloud without any systematic changes to the sequential structure of their thought processes. The fact that subjects must already possess the necessary skills for efficient verbalization of thoughts is consistent with extensive evidence on the acquisition of self-regulatory private speech during childhood and on the spontaneous vocalization of inner speech by adults, especially in noisy environments. (Ericsson 2003)

Ericsson and Simon's empirical work on their protocol analysis method has confirmed it as a valid experimental approach, which has become widely acknowledged among the general psychology community. So, although verbal thought is fallible, there is some reason to think it can be generally trusted in certain contexts.

As I see it, this leaves us with a significant need for more work in identifying and assessing the epistemic qualities of inner speech. Considering the current status of empirical data and conceptual distinctions that pertain to the epistemology of inner speech, as I have outlined above, it is clear that it does not admit of any singular characterization. I regard the acknowledgment of this point as progress, but it is really just a beginning, identifying a need for a broad range of comparative inquiry into the epistemic strengths and weaknesses of our diverse introspective capacities and the implications that they have in our lives. I hope that the work I have provided here may serve as a foundational framework for this sort of inquiry, such that our understanding, and misunderstanding, of ourselves through the incessant monologues we find in our minds can itself be more adequately understood.

Before ending, however, I want to offer some somewhat more speculative considerations that extend the ideas I have surveyed in this section. In regard to the fallibility of our inner speech, I am willing to take this a bit further and go so far as to suggest that inner speech can in some cases lead us into self-deceptive states where we engage in narrative rationalization of our own mental states in such a way that we mitigate or cover up aspects of ourselves that could otherwise be overtly apparent to

us. Once again, this leads into vast territory that I can only hazily outline. There is already a significant body of work on the possibility and cognitive mechanics of self-deception, to which I simply cannot do justice to here (for entry into the literature on self-deception see the articles in McLoughlin and Rorty 1988 and Dupuy 1998). Nevertheless, it is still worth considering here the possibility of self-deception through inner speech, which I will illustrate with an example. Imagine the case of a married man who meets an attractive woman co-worker. Upon talking to her, he finds that they get along quite well. She is very nice to him and shares many of his interests, and over a period of a few weeks they get to know each other in a very friendly manner. Through this interaction, the man develops feelings for the woman and finds himself thinking about her often. However, this also brings about a sense of fear in the man. He begins to feel threatened by the woman, in the sense that his involvement with her presents a danger to his marriage. At this point, he starts to have defensive thoughts that question the nature of the woman, telling himself things like "I think there's something devilish working in her. She seems deceitful. I don't think I really like her after all." In reality, the woman is quite nice and has no ill intentions towards the man. In fact, she has actually come to think of the man as a good friend. The man, however, through the construction of negative thoughts about the woman in his inner speech, eventually convinces himself that he does not like the woman. In this case, it seems reasonable to me to regard the man as self-deceived. In reality, he has established the beginnings of friendship with the woman. She is friendly to him, shares some of his interests, and on an intuitive, visceral level he is

drawn to interact with her. But, through confabulatory thoughts in inner speech, he convinces himself that he does not like her. He deceives himself about his feelings, replacing a budding friendship with thoughts of dislike among the propositions he utters through inner speech. With this being at least a conceivable situation, we can see that our inner speech could possibly lead us to believe false propositions about our own mental states that would otherwise be readily acknowledged. So, inner speech can be self-deceptive in some circumstances.

However, I do not think that the story ends there. In the case of the man above, we can imagine that over time he may actually grow to dislike the woman. As the man continues to engage in negative inner speech about the woman, his feelings of camaraderie and attraction will gradually subside, and he could even become less interested in the things they initially held in common. So, eventually, the beliefs that the man talked himself into may actually become true, such that he really does not like the woman and consistently behaves accordingly. With this extension of the story, we can see how what was first a case of self-deception through inner speech can become a case of self-fulfilling conceptualization of oneself.

In interpretation of this possibility, I suggest that the epistemic fallibility of inner speech, as it pertains to knowledge of our own minds, is not necessarily a straightforward case of constructing false beliefs. Now, as illustrated in the various examples I've discussed, we can clearly come to have false beliefs about ourselves, but to leave the impact of inner speech on self-knowledge at that would be to miss something. Through our inner speech, we can not only come to believe false things

about our own minds (or, of course, believe true things about our minds, when we get it right), we can also construct, shape, and impact our own mentality. Perhaps in a sense we at least partially create our own nature over time through the things we tell ourselves in inner speech. In other words, perhaps through weaving narrative stories and theories about ourselves we are to some extent creating ourselves. Some have made similar points in regard to understanding the nature of the self (i.e. Dennett 1989 and 1992), but I want to separate this point from the attempt to explain what the self is, per se. Whatever else we might be, or whatever else our selves may consist of, we do think about ourselves through language in inner speech, and in consideration of examples like that above, what we tell ourselves about ourselves can impact what is true of our own mental states. In short, we can not only know and fail to know our own minds through inner speech, but also at least partially determine what there is to know about our minds.

7.6 Conclusion

In this chapter, I have explained and defended the view that language, in the form of inner speech, is constitutive of conscious propositional thought. Stated succinctly, the idea is that we think in and with words. One central aspect of this view is that language brings objects of thought to conscious attention in the form of propositional language, enabling us to be consciously aware of our own propositional thoughts as they occur to us. Moreover, the language of inner speech serves a

recursive function in human thought, allowing us to consciously think about our own thoughts. I have disagreed with the view that this recursive function of language is the basis of consciousness itself, but defended the weakened view that it serves as the basis for conscious propositional thought and self-reflection upon such thought, which are, of course, important facets of our introspective abilities.

I have suggested that the self-knowledge enabled by inner speech consists of three distinct types. First, there is the phenomenal knowledge of having a conscious propositional thought, experienced as the utterances that occur in inner speech. Secondly, we obtain knowledge with propositional content through the use of language to conceptually represent aspects of our own minds in propositional form. Finally, we can further use language to conceptually represent our propositional thoughts themselves, thereby enabling recursive self-knowledge of our own propositional thoughts about our own minds.

Through further analysis of more specific epistemic qualities of inner speech, I have noted that the terrain here is quite varied, and in need of further inquiry. In some cases, our utterances in inner speech self-determine their own truth value, while in other cases we may come to have false beliefs about ourselves. Moreover, in some cases we may even deceive ourselves through inner speech, which could in turn have a self-fulfilling impact on our mental states, such that we actually take on the mental characteristics that we initially deceived ourselves about. Considering these things, further empirical inquiry and theoretical analysis of the epistemic qualities of inner speech are clearly called for, and could provide further substantial insight into how

we can know, fail to know, and perhaps also create our own mental states through our internal monologues.

CHAPTER 8 – CONCLUSION: THE NATURE, SCOPE, AND LIMITS OF INTROSPECTION

In this dissertation I have provided a new framework for understanding introspection. Rather than being a single, unified faculty of the mind, as is commonly thought or simply assumed, introspection is a heterogeneous collection of different kinds of first-person self-knowledge, with a variety of different epistemic qualities. By highlighting and outlining the broad and diverse nature of introspection, I hope to have improved our general understanding of how we know about our own minds, along with the nature and extent of such knowledge. I believe the account I have given provides a more accurate portrayal of introspection, describing the various ways in which we know, and fail to know, our own mental states in a more nuanced manner than has been previously available. In closing, I would like to call attention to what I believe to be the most significant aspects of this new account of introspection.

First of all, I have shown that the common understanding of introspection as inner perception is not literally accurate and involves the projection of perceptual metaphor onto our knowledge and understanding of our own minds. Among the most salient reasons for this rejection of introspection as a kind of perception are the fact that there are no physiological processes that can be identified with mental perception, a lack of mental objects with perceptually identifiable properties, and the inability of perception to account for the recursive nature of introspection. Taken

together, these reasons provide a strong case against the literal interpretation of introspection as a kind of perception, and opens up the concept of introspection for reevaluation, such that a more accurate framework for understanding the nature of introspection can be developed. I provide this much-needed framework in this dissertation, identifying the major cognitive factors and processes that underlie our ability to know our own minds.

Perhaps the most basic feature of our ability to know our own minds is the experiential self-knowledge that is intrinsic to being in a conscious state. By the very fact that our minds are at least partially constituted by conscious states, we acquire a kind of non-propositional knowledge such that we know what it is like, phenomenally speaking, to be in those conscious states. This kind of knowledge is unique to our first-person experience of our own mentality, and is quite different in nature from the prototypical processes that usually inform our understanding of knowledge (i.e. perception and rational inference). The unique nature of this kind of knowledge accounts for some common intuitions regarding introspection, such as the Cartesian idea that we infallibly know our own minds. However, it must also be acknowledged that this aspect of introspective self-knowledge is somewhat limited, pertaining only to the phenomenal experience of a conscious state. This experiential knowledge, contrary to what many have inferred, does not tell us anything propositional about the nature of the mental states that we experience, such as whether or not they can be understood in terms of physical processes. To obtain any propositional understanding

of our own minds, whereby we can make propositional claims about their nature and their properties, we must resort to additional cognitive processes.

This brings us into the domain of representation, conceptualization, and attention. Although other cognitive processes play roles in our introspective abilities (memory, for example), these three fundamental aspects of human cognition are particularly important in understanding how we are capable of knowing, and failing to know, our own minds. The representational capacities of the mind enable us to represent our own mental states, such that our own mental states can become the objects of thought and reasoning, much as representing aspects of the external world enables us to think and reason about them. Overlapping with this is the ability to form concepts, which allows us to place the various phenomena we experience into coherent theoretical frameworks that makes sense of them. So, just as we utilize the concept of gravity to understand certain events in the world around us, we can utilize concepts of mental phenomena (such as the concepts of pain, emotion, belief, etc) to understand what is happening in our own minds. In addition, attention enables us to guide these cognitive resources towards various phenomena, such that we can actively reflect upon and think about our own mental lives. Taken together, these cognitive processes provide a powerful explanatory framework that can account for many of our introspective abilities and accommodate them within cognitive processes that we already have some understanding of. Although these processes provide the basis for significant aspects of introspective self-knowledge, it is crucial to realize that they also open the door for epistemic errors in introspection. Intrinsic to the

mediated nature of representation and conceptualization are the possibilities of misrepresenting and applying mistaken concepts to aspects of our own mental lives, such that we can not only know, but also fail to know, our own minds.

Related to this is the fact that our understanding of our own beliefs and desires is mediated by a theoretical structure commonly known as folk psychology. While it may seem that we immediately know what we ourselves believe and desire, I have shown how understanding these aspects of our own minds involves the application of abstract concepts that may or may not accurately reflect the actual contents of our own minds. To support this somewhat counterintuitive idea, I have described several different avenues of empirical research (from experimental neuroscience, social psychology, and developmental psychology) that indicate that our understanding of our own folk psychological states is in fact mediated in this manner. So, just as we inferentially apply folk psychological concepts when understanding the behavior of others, we likewise use this understanding to infer from our first-person experiences the presence of folk psychological states in our own minds.

Overlapping with the various cognitive processes that enable our ability to introspect is the phenomenon of inner speech. In interpretation of this phenomenon, I have suggested that the language of inner speech is a vehicle for conscious propositional thought, with which we are consciously aware of our own thoughts and, moreover, recursively engage in conscious thought about our own thoughts. It is important to note that this is somewhat weaker than many other cognitive conceptions of language. I do not claim that natural language is the basis for thought, but rather

only one cognitive device, among possibly many others, with which we think. Additionally, I do not claim that natural language is the basis for consciousness in general, but rather only the basis for conscious awareness of our own propositional thoughts. Considering these somewhat narrowed claims regarding the cognitive abilities enabled by natural language, we can plausibly regard inner speech as an important facet of introspective cognition. From this basis, I go on to discuss multiple epistemic properties of inner speech. In some cases, what we utter to ourselves in inner speech constitutes a self-constitutive truth, such that the utterance is rendered true by its very occurrence (such as the thought "I am thinking that I am awake right now"). However, in many other cases the utterances of inner speech are significantly fallible, and leave open the possibility of making erroneous statements about our own minds. In this regard, inner speech overlaps with the contingent fallibility of the representational and conceptual processes mentioned earlier. At the far end of this spectrum lies the capacity to deceive ourselves, such that what we utter in inner speech can at times rationalize and cover up aspects of ourselves that would otherwise be readily apparent to us. Over time, however, such self-deception can sometimes become self-fulfilling, when what we initially deceive ourselves about becomes integrated into our thoughts and behavior in such a way that they become true of us. So, even within the domain of inner speech, we can see that epistemic characteristics of introspection are quite varied, providing further support for my overall perspective of introspection as a heterogeneous phenomenon.

This concludes my dissertation. I hope that the framework I have provided here concerning the nature, scope, and limits of our introspective abilities will serve to clarify and improve understanding of the diverse domain of first-person self-knowledge commonly known as introspection.

BIBLIOGRAPHY

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

Armstrong, D. M. 1997. What is consciousness? In Block, Flanagan, and Güzeldere, eds. 1997: 721-728.

Baars, B. J. 1997. *In the Theater of Consciousness: The Workspace of the Mind*. New York: Oxford University Press.

Bar-On, D. 2004. *Speaking My Mind: Expression and Self-Knowledge*. New York: Oxford University Press.

Bermúdez J. L., A. Marcel, and N. Eilan, eds. 1995. *The Body and the Self*. Cambridge: MIT Press.

Bermúdez, J. L. 1998. *The Paradox of Self-Consciousness*. Cambridge: MIT Press.

Block, N., O. Flanagan, and G. Güzeldere, eds. 1997. *The Nature of Consciousness:Philosophical Debates*. Cambridge: MIT Press.

Botterill, G. and P. Carruthers. 1999. *The Philosophy of Psychology*. Cambridge: Cambridge University Press.

Byrne, R. W. and A. Whiten. 1988. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans.* Oxford: Clarendon Press.

Carruthers, P. and P. K. Smith, eds. 1996. *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Carruthers, P. 1996. *Language, Thought and Consciousness: An Essay in Philosophical Psychology.* Cambridge: Cambridge University Press.

Carruthers, P. 1998. Conscious Thinking: Language or Elimination? *Mind and Language* 13(4): 457-476.

Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Chambres, P., M. Izaute, and P. Marescaux, eds. 2002. *Metacognition: Process, Function, and Use*. Boston: Kluwer Academic Publishers.

Churchland, P. M. 1981. Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy* 78 (2): 67-90.

Churchland, P. M. 1984. *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind*. Cambridge: MIT Press.

Churchland, P.M. 1985. Reduction, Qualia, and the Direct Introspection of Brain States. *Journal of Philosophy* 82 (1): 8-28.

Churchland, P. S. 2002. *Brain-Wise: Studies in Neurophilosophy*. Cambridge: MIT Press.

Conee, E. 1994. Phenomenal Knowledge. *Australasian Journal of Philosophy*. 72 (2): 136-150.

Crick, F. 1994. *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Touchstone.

Crumley, J. S. 2000. *Problems in Mind: Readings in Contemporary Philosophy of Mind*. Mountain View, CA: Mayfield Publishing Company.

Csikszentmihalyi. M. 1997. *Finding Flow: The Psychology of Engagement with Everyday Life*. New York: Basic Books.

Cummins, D. D. and C. Allen. 1998. *The Evolution of Mind*. New York: Oxford University Press.

Damasio, A. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace and Company.

Danziger, K. 1980. The History of Introspection Reconsidered. *Journal of the History of the Behavioral Sciences* 16: 241-262.

DeCharms, C. 1998. *Two Views of Mind: Abhidharma and Brain Science*. Ithica, NY: Snow Lion Publications.

Dennett, D. C. 1989. The Origins of Selves. *Cogito* 3: 163-173.

Dennett, D. C. 1991. *Consciousness Explained*. Boston: Little, Brown, and Co.

Dennett, D. C. 1992. *The Self as a Center of Narrative Gravity*. In Kessel, F., P. Cole, and D. Johnson, eds. *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Erlbaum.

Dennett, D. C. 1998. *Brainchildren: Essays on Designing Minds*. Cambridge: MIT Press.

Dretske, F. 1995. *Naturalizing the Mind*. Cambridge: MIT Press.

Dupuy, J., ed. 1998. *Self-Deception and Paradoxes of Rationality*. Stanford: Center for the Study of Language and Information Publications.

Ericsson, K. A. and H. A. Simon. 1984. *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.

Ericsson, K. A. 2003. Valid and Non-Reactive Verbalization of Thoughts During Performance of Tasks. In Jack and Roepstorff, eds. 2003: 1-18.

Falvey, K. 2000. The Basis of First-Person Authority. *Philosophical Topics* 28 (2): 69-99.

Farah, M. J. 2000. *The Cognitive Neuroscience of Vision*. Malden, MA: Blackwell Publishers Inc.

Finkelstein, D. H. 2003. *Expression and the Inner*. Cambridge, MA: Harvard University Press.

Fodor, J. A. 1987. *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J. A. 1998. *Concepts: Where Cognitive Science Went Wrong.* New York: Oxford University Press.

Fodor, J. A. 2001. Language, Thought and Compositionality. *Mind and Language* 16 (1): 1-15.

Fuster, J. M. 2002. Physiology of Executive Functions: The Perception-Action Cycle. In Stuss and Knight, eds. 2002: 96-108.

Gaulin, S. J. C. and D. H. McBurney. 2001. *Psychology: An Evolutionary Approach.* Upper Saddle River, NJ: Prentice-Hall, Inc.

Gazzaniga, M. 1998. *The Mind's Past*. Los Angeles, CA: University of California Press.

Goldman, A. 1993. The psychology of folk psychology. *Behavioral and Brain Sciences* 16: 15-28.

Gopnik, A. 1993. How we know our own minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences* 16: 1-14.

Gopnik, A. and A. Meltzoff. 1997. *Words, Thoughts, and Theories*. Cambridge: MIT Press.

Gregory, R. L. 1998. *Eye and Brain: The Psychology of Seeing*. 5th edition. New York: Oxford University Press.

Güzeldere, G. 1997. Is Consciousness the Perception of What Passes in One's Own Mind? In Block, Flanagan, and Güzeldere, eds. 1997: 789-806.

Hardcastle, V. G. 1998. The puzzle of attention, the importance of metaphors. *Philosophical Psychology* 11 (3): 331-351.

Heil, J. 1988. Talk and Thought. *Philosophical Papers* 17 (3): 153-170.

Heyes, C. M. 1998. Theory of mind in nonhuman primates. *Behavioral and Brain Sciences* 21: 101-114.

Hofstadter, D. R. 1979. *Gödel, Escher, Bach: an Eternal Golden Braid*. New York: Basic Books.

Humphrey, N. 1986. *The Inner Eye*. New York: Oxford University Press.

Itti, L., G. Rees, and J. K. Tsotsos, eds. 2005. *Neurobiology of Attention*. Burlington, MA: Elsevier Academic Press.

Jackendoff, R. 1996. How language helps us think. *Pragmatics and Cognition* 4 (1): 1-34.

Jack, A. and A. Roepstorff, eds. 2003. *Trusting the Subject?: The Use of Introspective Evidence in Cognitive Science Vol. 1*. Charlottesville, VA: Imprint Academic.

Jackson, F. 1982. Epiphenomenal Qualia. *The Philosophical Quarterly* 32: 127-136. Reprinted in Crumley, ed. 2000: 556-563.

Jackson, F. 1986. What Mary Didn't Know. *The Journal of Philosophy* 83 (5): 291-295. Reprinted in Crumley, ed. 2000: 577-580.

Jacob, P. and M. Jeannerod. 2003. *Ways of Seeing: The Scope and Limits of Visual Cognition*. New York: Oxford University Press.

James, W. 1890. *The Principles of Psychology*. http://psychclassics.yorku.ca/James/Principles/ (accessed March 17, 2004).

Kant, I. 1985. *Critique of Pure Reason*. Trans. N. K. Smith. http://www.arts.cuhk.edu.hk/Philosophy/Kant/cpr/ (accessed March 17, 2004). First Published in 1781.

Kim, J. 1998. *Philosophy of Mind*. Boulder, CO: Westview Press.

Kolak, D. and R. Martin 1991. *Self and Identity: Contemporary Philosophical Issues*. New York: MacMillan Publishing Company.

Kornblith, H. 1998. What is it like to be me? *Australasian Journal of Philosophy* 76 (1): 48-60.

Kornblith, H. 2002. *Knowledge and its Place in Nature*. New York: Oxford University Press.

LaBerge, D. 1995. *Attentional Processing: The Brain's Art of Mindfulness.* Cambridge: Harvard University Press.

Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, G. and M. Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Lillard, A. 1998. Ethnopsychologies: Cultural Variations in Theories of Mind. *Psychological Bulletin* (American Psychological Association). 123 (1): 3-32.

Locke, J. 1975. *An Essay Concerning Human Understanding.* Ed. P. H. Nidditch. Oxford: Clarendon Press. First published in 1689.

Lycan, W. G. 1987. *Consciousness*. Cambridge: MIT Press.

Lycan, W. G. 1996. *Consciousness and Experience*. Cambridge: MIT Press.

Lyons, W. 1986. *The Disappearance of Introspection.* Cambridge: MIT Press.

Machery, E. 2005. You Don't Know How You Think: Introspection and Language of Thought. *British Journal of the Philosophy of Science* 56: 469 – 485.

Margolis, E. and S. Laurence, eds. 1999. *Concepts: Core Readings*. Cambridge: MIT Press.

Mazzoni, G. and T. O. Nelson, eds. 1998. *Metacognition and Cognitive Neuropsychology: Monitoring and Control Processes*. Mahwah, NJ, Lawrence Erlbaum Associates.

McLaughlin, B. P and A. O. Rorty, eds. 1988. *Perspectives on Self-Deception*. Berkeley: University of California Press.

Meissner, C. A. and A. Memon. 2002. Verbal Overshadowing: A Special Issue Exploring Theoretical and Applied Issues. *Applied Cognitive Psychology* 16: 869-872.

Metcalfe, J. and A. P. Shimamura, eds. 1994. *Metacognition: Knowing about Knowing*. Cambridge, MA: MIT Press.

Moran, R. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press.

Morton, A. 1980. *Frames of Mind: Constraints on the Common-Sense Conception of the Mental*. Oxford: Clarendon Press.

Musgrave, A. 1993. *Common Sense, Science, and Scepticism*. New York: Cambridge University Press.

Nadel, L., ed. 2003. *Encyclopedia of Cognitive Science*. New York: Nature Publishing Group.

Nagel, T. 1974. What is it like to be a bat? *The Philosophical Review* 83 (4): 435-50. Reprinted in Block, Flanagan, and Güzeldere, eds. 1997: 519-527.

Neisser, U., ed. 1993. *The Perceived Self: Ecological and interpersonal sources of self-knowledge*. Cambridge: Cambridge University Press.

Nichols, S. and S. Stich. 2003. Reading One's Own Mind: A Cognitive Theory of Self-Awareness. http://ruccs.rutgers.edu/ArchiveFolder/ Research%20Group /Publications/Room/room.html (accessed November 21, 2003).

Nida-Rümelin, M. 2002. Qualia: The Knowledge Argument. *The Stanford Encyclopedia of Philosophy (Fall 2002 edition)*. E. N. Zalta, ed. http://plato.stanford.edu/archives/fall2002/entries/qualia-knowledge/ (accessed December 15, 2002).

Nisbett, R. and T. D. Wilson. 1977. Telling More than We Can Know: Verbal Reports on Mental Processes. *Psychological Review* 84: 231-259.

O'Shaughnessy, B. 1995. Proprioception and the Body Image. In Bermúdez, Marcel, and Eilan, eds. 1995: 175-203.

Palmer, J. A. and L. K. Palmer. 2002. *Evolutionary Psychology: The Ultimate Origins of Human Behavior*. Boston: Allyn and Bacon.

Pashler, H. E. 1998. *The Psychology of Attention*. Cambridge, MA: MIT Press.

Penfield, W. and T. Rasmussen. 1950. *The Cerebral Cortex of Man: A Clinical Study of Localization of Function*. New York: MacMillan.

Picton, T. W., C. Alain, and A. McIntosh. 2002. The Theatre of the Mind: Physiological Studies of the Human Frontal Lobes. In Stuss and Knight, eds. 2002: 109-126.

Pomerantz, J. R. 2003. Perception: Overview. In Nadel, ed. 2003: 527-537.

Powers, J. 1995. *Introduction to Tibetan Buddhism*. Ithica, NY: Snow Lion Publications.

Prabhavananda, S. and F. Manchester., trans. 1975. *The Upanishads: Breath of the Eternal*. New York: Mentor Books.

Premack, D. and G. Woodruff. 1978. Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences* 4: 515-526.

Ramachandran, V. S. and S. Blakeslee. 1998. *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. New York: HarperCollins Publishers.

Ristau, C. A. 1998. Cognitive Ethology: The Minds of Children and Animals. in Cummins and Allen, eds. 1998: 127-159.

Russell, B. 1910. Knowledge by Acquaintance and Knowledge by Description. *Aristotelian Society Proceedings* 11: 108-128.

Russell, B. 1956. Mind and Matter. in *Portraits from Memory*. Nottingham: Spokesman.

Sacks, O. 1970. *The Man Who Mistook His Wife for a Hat*. New York: Touchstone.

Schooler, J. W. and T. Y. Engstler-Schooler. 1990. Verbal overshadowing of visual memories: some things are better left unsaid. *Cognitive Psychology* 22: 36-71.

Searle, J. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Sellars, W. 1956. Empiricism and the philosophy of mind. *Minnesota studies in the philosophy of science* 1: 253-329. http://www.ditext.com/sellars/epm.html (Accessed November 21, 2003).

Shoemaker, S. 1994a. Self-Knowledge and "Inner Sense": Lecture I: The Object Perception Model. *Philosophy and Phenomenological Research* 54 (2): 249-269.

Shoemaker, S. 1994b. Self-Knowledge and "Inner Sense": Lecture II: The Broad Perceptual Model. *Philosophy and Phenomenological Research* 54 (2): 271-290.

Shoemaker, S. 1994c. Self-Knowledge and "Inner Sense": Lecture III: The Phenomenal Character of Experience. *Philosophy and Phenomenological Research* 54 (2): 291-314.

Silvia, P. J., and G. H. E. Gendolla. 2001. On Introspection and Self-Perception: Does Self-Focused Attention Enable Accurate Self-Knowledge? *Review of General Psychology* 5 (3): 241-269.

Sperber, D., ed. 2000. *Metarepresentations: A Multidisciplinary Perspective*. New York: Oxford University Press.

Sterelny, K. 1998. Intentional Agency and the Metarepresentation Hypothesis. *Mind and Language* 13 (1): 11-28.

Stich, S. and Nichols, S. 1998. Theory theory to the max: A critical notice of Gopnik & Meltzoff's *Words, thoughts, and theories. Mind & Language*, 13: 421-449. http://ruccs.rutgers.edu/ArchiveFolder/Research%20Group/Publications/g&m/G&M.html (accessed November 21, 2003).

Stuss, D. T. and R. T. Knight, eds. 2002. *Principles of Frontal Lobe Function*. New York: Oxford University Press.

Taylor, S. and J. Brown. 1988. Illusions and Well-Being: A Social Psychological Perspective on Mental Health. *Psychological Bulletin* 103: 193-210.

Thera, N. 1962. *The Heart of Buddhist Meditation*. New York: Samuel Weiser, Inc.

Wallace, B. A. 2000. *The Taboo of Subjectivity: Toward a New Science of Consciousness.* New York: Oxford University Press.

Williams, B. 1970. The Self and the Future. *The Philosophical Review* 79 (2): 161-180.

Wilson, T. D. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: The Belknap Press of Harvard University Press.

Zahavi, D. 2002. First-person thoughts and embodied self–awareness: Some reflections on the relation between recent analytical philosophy and phenomenology. *Phenomenology and the Cognitive Sciences* 1: 7-26.