

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

IDENTIFYING THE FIRST PERSON

by

ROBLIN R. MEEKS

A dissertation submitted to the Graduate Faculty in
Philosophy in partial fulfillment of the requirements for
the degree of Doctor of Philosophy, The City University of
New York

2003

UMI Number: 3103147

Copyright 2003 by
Meeks, Roblin Roy

All rights reserved.

UMI[®]

UMI Microform 3103147

Copyright 2003 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

©2003

ROBLIN R. MEEKS

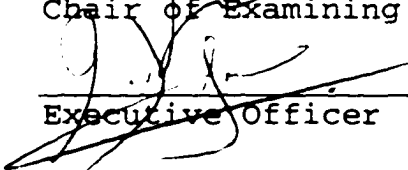
All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

9/10/03
Date

9/10/03
Date


Chair of Examining Committee


Executive Officer

Stephen Grover

Michael Levin

David M. Rosenthal
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

IDENTIFYING THE FIRST PERSON

by

Roblin R. Meeks

Advisor: Professor David M. Rosenthal

Wide agreement exists that self-ascriptions that one would express with the first-person pronoun differ in kind from those one would express with other self-designating expressions such as proper names and definite descriptions. At least some first-person self-ascriptions, many argue, are nonaccidental--that is, they involve no self-identification, and hence in making them one cannot accidentally misidentify the subject of the ascription. I examine the support for this claim throughout the literature, paying particular attention to Sydney Shoemaker's proposal that self-ascriptions are nonaccidental in virtue of being immune to error through misidentification relative to the first-person pronoun. According to Shoemaker, such immunity results from the special way in which one is introspectively aware of the psychological property or state ascribed, a way that leaves no room for questions to arise as to whether oneself is its bearer. I contend that though it may seem from the point

of view of consciousness that we are directly and immediately aware of the states of our bodies and minds as our own, both theoretical and empirical considerations strongly suggests that we have no such direct awareness. Proprioception and introspection prove in the end to be better described as types of informed, conscious self-interpretation. Taking inspiration from Dennett, Rosenthal, and Nozick, I offer the naïve proposal as an alternative that explains all self-ascriptions in terms of one's relying upon a battery of commonsense self-specifying beliefs to interpret both which state or property one has and who has it. As a result, first-person self-ascriptions differ from others only in degree and not in kind, and self-misidentification always remains a possibility--even when self-ascribing properties with the first-person pronoun.

Acknowledgements

This work would not be what it is without the help and generosity of a host of people. In particular, I would like to thank fellow students and friends Jennifer Fisher, Andres Paez, Douglas Meehan, Josh Weisberg, Jared Blank, Bill Seeley, Carrie Figdor, Russell Marcus, and Eric Hetherington. I have enjoyed and benefited enormously from your wit and wisdom. The Graduate Center would not be as good as it is without you.

A number of professors also deserve credit. Michael Levin was truly indefatigable in reading and commenting upon drafts throughout the writing progress. The other members of my committee, Jonathan Adler, John Greenwood, and Stephen Grover, each gave invaluable input and suggestions for further research. As a visiting professor, Jaegwon Kim piqued my interest in the topic of self-ascription, and Simon Blackburn helped me to unravel some confusion. Pete Mandik was also always quick with astute and welcome criticism. I was also privileged to have at my disposal in the final stages of writing the acumen and encouragement of Frederique de Vignemont, who never failed to demonstrate how philosophy and psychology can truly complement one another. And my advisor David Rosenthal

incessantly challenged me to think more deeply and to produce work that I could be proud of long after my student days have passed.

Parts of this thesis were presented at a variety of conferences, including the Pacific APA, the Association for the Scientific Study of Consciousness, the Society for Philosophy and Psychology, the European Society for Philosophy and Psychology, the New Jersey Regional Philosophical Association, and several Cognitive Science Symposiums. I received many useful comments and suggestions from both commentators and audiences.

Finally, I would above all like to thank my family. My mother and father, Nancy and Byron Meeks, never doubted my abilities for a moment, even when doubts of my own arose. Rowdy Meeks, my brother, always helped me to put this project in perspective and was always willing to listen to my laments. The love, grace, and patience of my wife Yen cannot be overstated. She made the years--many of them difficult--tolerable and navigable; without her neither I nor what follows would exist. My son Mercer likewise deserves mention. Though he surprised us all by arriving early, he proves just how fascinating small selves are.

IDENTIFYING THE FIRST PERSON

Table of Contents

CHAPTER I. INTRODUCTION AND FIRST IMPRESSIONS	1
CHAPTER II. THE PROBLEM OF SELF-IDENTIFICATION	16
§1. Introduction.....	16
§2. Framing the Problem: Characterizing First-Person Thought.....	19
§3. Chisholm: The Irreducibility Argument and Fundamental Intentionality.....	27
§4. Revisiting Chisholm's Conclusion.....	30
§5. An Alternative: The Naïve Proposal of Self- Identification Introduced.....	43
§6. Objections to NP.....	52
§7. Conclusion.....	57
CHAPTER III. ASSESSING ASCRIPTIVE ACCIDENTS: GUARANTEED REFERENCE AND IMMUNITY TO ERROR THROUGH MISIDENTIFICATION	59
§1. Introduction.....	59
§2. Wittgenstein's 'I': First-Person Pronoun Used as Subject and Object.....	64
§3. Nonaccidental Self-Ascriptions and Guaranteed Self- Reference.....	84
§4. Immunity to Error Through Misidentification Relative to the First-Person Pronoun.....	106
§5. Making Sense of IEM.....	120
§6. Conclusion.....	131
CHAPTER IV. THE BODY AND ITS OWNER: IMMUNITY AND THE SELF- ASCRPTION OF BODILY PROPERTIES	134
§1. Introduction.....	134
§2. Understanding the Regress: Beginnings and Endings	137
§3. Evans: Proprioception, Bodily Self-Ascriptions, and IEM.....	147
§4. Bermúdez: Nonconceptual Bodily Self-Ascriptions and IEM.....	162
§5. Errors of Bodily Identification.....	177
§6. NP and Bodily Self-Ascriptions: The Body Image...	184
§7. Conclusion.....	194
CHAPTER V. MISIDENTIFYING THE FIRST PERSON	198
§1. Introduction.....	198
§2. Consciousness, Introspection, and Self-Interpretation	202
§3. Who's Interpreting Whom: Self-Interpretation, Self- Identification, and Unity of Self.....	225

§4. The Not-So-Naïve-Proposal: A Review.....	252
§5. Answering the Objections: The Misconception Problem and the Amnesia Problem.....	260
§6. The Asymmetry Problem.....	264
§7. Conclusion: Avenues of Exploration.....	266
BIBLIOGRAPHY	270

Chapter I. Introduction and First Impressions

"I: I is the first letter of the alphabet, the first word of the language, the first thought of the mind, the first object of the affections. In grammar it is a pronoun of the first person and singular number. Its plural is said to be We, but how there can be more than one myself is doubtless clearer to the grammarians that it is to the author of this incomparable dictionary. Conception of two myselfs is difficult, but fine."

--Ambrose Bierce, The Devil's Dictionary

Bierce's clever definition nicely captures the intuitive sense we have of ourselves as thinkers. The first-person is indeed aptly named: Intuitively, thoughts about ourselves that we express with it or involve its mental analogue seem to have a special status. The first person seems to lie hidden in some way from the second and third, glimpsed by them only second hand in speech or behavior. And asking how one thinks about oneself sounds at best like a joke or a mistake--nothing could be more simple. As Beirce suggests, it's almost impossible for one not to be, as it were, selfish in thought--not to take oneself as "the first object of the affections."

These intuitions about the first person give rise to a host of difficult philosophical puzzles. How does one self-ascribe states and properties, both mental and

physical? Intuition suggests that one has privileged access to one's mental life, and hence one need only ask oneself what one is thinking. But is that intuition accurate, and if so, how can we understand this access? Is our thinking about ourselves fundamentally different from our thinking about others? If we do enjoy some form of ascriptive privilege, are one's mental states and properties therefore in principle removed from others? Does consciousness include an ineliminable first-person element such that no "objective", third-person scientific theory can hope to explain the phenomenon fully and satisfactorily?

Historically, philosophers have tended to accept first-person privileged access as a fact about our mental lives, ultimately putting it to good theoretical use. Perhaps the best example is that of Descartes, who argues that only a first-person judgment can defeat radical philosophical skepticism and provide the foundation upon which to build secure and certain scientific knowledge. Moreover, Descartes held that the mind is transparent to itself, and one cannot be in a mental state without also being aware that one is in that state.

The traditional problem of other minds and its traditional solution provides another illustration of the

philosophical appeal of the first person. Russell, for one, followed Mill in arguing that one only ascribes states to others indirectly by first observing the causal connections between one's own states and one's behavior. One can then infer that similar observed behavior must be the result of similar unobservable mental states. Privileged access may mean that no one else can directly observe one's mental states, but those who endorse the argument from analogy believe that this privilege itself also offers a solution to the problems it creates.

Many difficulties have been raised for the argument from analogy, of course. Critics are quick to point out, for example, that relying upon a single case as the basis for a large number and range of conclusions is a poor inductive strategy. More interesting, however, are the challenges to privilege itself. One's assertions about one's mental states were thought to be if not infallible then at least incorrigible. But in the wake of powerful arguments like that given by Wittgenstein and his followers against the possibility of a private language, few philosophers remain convinced that such absolute privilege exists. Most now accept that introspection provides reliable though fallible access to one's mental states.

One may be an authority with regard to one's mind--but not an unassailable one.

My primary interest in the dissertation is not the problem of other minds or the privileged access that precipitates it. I am puzzled by another type of Cartesian privilege that remains widely endorsed, however, one lurking in Beirce's assertion that first-person thoughts are the first thoughts of the mind. Again Descartes' Meditations provides a ready example of this notion. It's important for Descartes that first philosophy is in the first person--it would not suffice for him to conclude via the cogito that some subject or other exists (apart from God). For the existence of an external world and those who might inhabit it can be doubted. Descartes does not worry that the evil genius might deceive him into falsely thinking that the doubting subject and the subject doubted are one and the same. Indeed the cogito only succeeds if the doubter doubts his own existence, which is why opponents such as Lichtenberg and Nietzsche object at this point that Descartes oversteps his argument by framing his conclusion in first-person terms.

This further type of privilege concerns the identification of the ascribing subject. However fallible one might be about an ascribed state or property, few

accept that one can be mistaken about the identity of the ascriber when that state or property is ascribed with the first-person pronoun or its mental analogue. Though one can be wrong about which state one is in, it seems almost a truism that one cannot be wrong about who is in that state when one is introspectively aware of it.

So widely accepted is this type of first-person privilege or a variation of it that it often makes for strange bedfellows. Perhaps the most influential modern proponent of it is Wittgenstein, who contends that the first-person pronoun has a special use--what he calls 'I' used as subject--for which it is strictly nonsensical to question who is using the term. When 'I' is used in this way--as in, e.g., "I have toothache"--Wittgenstein writes that "there is no question of recognizing a person" and so no room remains for the possibility of misrecognition or misidentification (1969, p. 67). To avoid the outright Cartesianism intimated by this privilege, Wittgenstein claims that speech acts involving 'I' used as subject do not serve to pick out or to indicate a speaker. Surface grammar notwithstanding, they differ from other uses of the first-person pronoun--'I' used as object--in that they are tantamount to groans.

Sydney Shoemaker, much influenced by Wittgensteinian, in effect lays bare the Cartesian nature of this privilege. He agrees with Wittgenstein that certain first-person self-ascriptions enjoy a special freedom from error with regard to identifying the subject, but he wisely does not want to deny that 'I' is a referring expression even in these cases. The resulting position requires that he argue for what he calls weak self-intimation--namely the Cartesian notion that when one is in a particular mental state one will believe that one is in that state.

Still, this type of privilege proves rather difficult to understand. Intuitively, it does seem to us in introspection that we are directly and immediately aware of both the state introspected and oneself as a single, unified subject who is in that state. To introspect a thought appears prima facie inseparable from grasping that it is to oneself who believes or desires such-and-such. It doesn't seem to one that one first observes with one's "inner eye" some subject or other who believes such-and-such and then concludes that that believing subject is in fact oneself. No room for error regarding the introspecting subject seems to exist as Wittgenstein claims.

Part of the difficulty in understanding this notion of privilege lies in clarifying exactly what kind and degree of error of misidentification supposedly cannot occur. We have fairly strong intuitions regarding the possibility of errors in these cases, but our equally strong intuitions about privileged access to our states were eventually vitiated. Accordingly, my primary concern in the dissertation is the notion of self-identification and its role in self-ascription.

Self-identification should not be conflated with self-reference, however. I have little to add to the rich and extensive literature addressing the semantic properties of the first-person pronoun. In the dissertation I rely upon that literature mainly to illustrate the important distinction between self-reference and self-identification and their associated errors. One identifies oneself by possessing and expressing, in behavior and speech, beliefs about the properties one instantiates. Errors of misidentification occur when one refers to an object or individual while being disposed to deny that one is doing so. Similarly, self-misidentification occurs when one both refers to oneself and is disposed to deny that one is doing so, and this occurs most often when one self-ascribes a state or property with a self-designating expression other

than the first-person pronoun. One instantiates myriad more properties than one believes one possesses, and one no doubt believes that one possesses properties that one doesn't in fact instantiate.

The importance of these types of error cannot be overstated since they help to reveal the nature of the privilege in question. Intuition provides only a rough idea of what errors can occur. Accordingly, consonant with work in cognitive science, in the discussion to follow I rely quite heavily on empirical work primarily from cognitive psychology to expand and to sharpen the catalogue of ascriptive errors that one is subject to. We learn that empirical data provides compelling evidence of a wide variety of ascriptive errors that we are prone to even in the normal case. And the more extreme cases of abnormal mental functioning give us further insight into how we think about ourselves. Beirce correctly points out that "conception of two myselfs is difficult," for example, but clinical work regarding psychopathologies such as dissociative identity disorder and thought insertion helps us to achieve a clearer sense of how the mind views and associates its own states. In particular, we see that the traditional, intuitive picture of introspection as a direct

and unmediated awareness of one's own mental states and properties does not survive scrutiny.

I believe that this and similar work demonstrates that no such privilege regarding the identification of the introspecting subject actually obtains and that accounts of self-ascription including such privilege cannot accommodate the range of ascriptive errors elucidated both within and beyond the lab. The challenge, then, is to offer an alternative account of self-ascription that is sensitive to these errors even at the expense of intuitions we may have about them. Just as privileged access with regard to one's own mental states has succumbed to theoretical pressures, I believe that this privilege with regard to the identification of the ascribing subject will also ultimately not survive our best theory.

The dissertation is divided into four remaining chapters as follows.

I devote Chapter II to laying out the specific problem of self-identification in detail as it relates to our understanding of the self-ascription of properties and states. I borrow Roderick Chisholm's formulation of the problem in terms of the entailment relations (and important lack thereof) between de re, de dicto, and de se oratio obliqua belief ascriptions. Chisholm explains these

entailments in terms of one's fundamental, unanalyzable ability to ascribe properties and to take oneself as an intentional object. His account does honor the intuition that self-ascriptions canonically expressed with the first-person pronoun--those I call nonaccidental--cannot fall prey to errors of self-misidentification. But it does not succeed equally well in accommodating the commonsense errors one seems open to when self-ascribing properties and states accidentally or with expressions involving self-designating terms other than the first-person pronoun such as proper names, definite descriptions, and demonstratives.

Taking inspiration from these commonsense errors, in Chapter II I also briefly and broadly introduce my preferred alternative account to be developed and defended at length--the naïve proposal or NP--according to which all self-ascriptions draw upon a battery of commonsense self-identifying beliefs about such properties as one's name, birth date, habits of mind, bodily appearance, etc. Its virtues and shortcomings are exactly opposite to those of the Chisholmian view--NP easily accounts for self-ascriptive accidents but seems unable to speak to our intuitions about the ways in which nonaccidental self-ascriptions appear free from error in an important respect. These two diametrically opposed views also provide a

framework for the discussion to follow, as NP is filled out and positioned between them. To succeed, an account of self-ascription must address the three problems I raise at the end of the chapter: the Misconception Problem, the Amnesia Objection, and the Asymmetry Problem. Together they challenge the necessity and sufficiency of the battery requirement at the heart of NP, including how the beliefs that constitute it fund nonaccidental self-ascriptions that seem to be made without drawing upon any evidence whatsoever.

The primary ground of difference between NP and a Chisholmian account is the latter's reliance upon the assumption that nonaccidental self-ascriptions differ in kind from accidental ones in that they involve no self-identification and hence cannot fall prey to self-misidentification. Accordingly, in Chapter III I attempt to sharpen and to motivate this assumption by looking to perhaps its most ardent defenders--namely Wittgenstein, G.E.M. Anscombe, and Sydney Shoemaker. Shoemaker ultimately provides the most compelling case in favor of the assumption, arguing that if all self-ascriptions involved self-identification then we would be doomed to a regress of identification that would preclude an explanation of how we are able to ascribe properties and

states to anything at all. Hence, he concludes that at least some self-ascriptions must be "immune to error through misidentification relative to the first-person pronoun" (hereafter 'IEM', Shoemaker 1968). Since only introspection could provide the kind of direct and unmediated awareness necessary to preclude errors of misidentification, Shoemaker contends that only introspective self-ascriptions of psychological states and properties prove to be IEM.

Shoemaker's regress argument raises a clear challenge to NP, since the requirement that one must possess a battery of self-specifying beliefs to self-ascribe properties even with the first-person pronoun seems to precipitate the regress in question. I therefore pursue the regress challenge to NP in Chapter IV, including investigating where the regress forces one to retreat to if successful. I examine arguments by Gareth Evans and José Luis Bermúdez who argue contrary to Shoemaker that certain bodily self-ascriptions are also IEM. However, we see that empirical data strongly suggests that self-ascriptions of bodily states and properties are indeed subject to errors of self-misidentification when one's beliefs about one's body are manipulated in the right way, perhaps by a clever psychologist.

If self-ascriptions of bodily properties aren't IEM, does Shoemaker's regress inevitably force us to accept the Cartesian-friendly alternative that only the self-ascription of introspected mental states and properties are IEM? Shoemaker himself favors this conclusion, of course, but in Chapter V I argue that even these ascriptions are subject to errors of misidentification. Here again, empirical evidence suggests that introspection amounts to a form of conscious self-interpretation. One ascribes states in virtue of taking the intentional stance towards oneself, predicting and explaining one's own states and behavior in light of one's battery of self-identifying belief, and as a result one can be mistaken about whether one is actually in the state that one appears introspectively aware of. Building upon this data and work by Rosenthal, Dennett, and Nozick, I further argue that even in the normal case interpretation also extends to the ascribing self. That is, just as one is not aware of an "introspected" state in a direct and unmediated way that could lead to IEM self-ascriptions, one is not directly aware of the introspecting self. The battery of self-identifying beliefs that provide the basis for introspective interpretation also give rise to an interpreted self. Interpreting which state one is in is to a certain degree inseparable from positing the self

that is interpreted. The regress argument can therefore be disarmed, for seen in this way an understanding of self-ascription need not require accepting the basic premise that ascribers have a primitive, privileged awareness of themselves.

The fully fleshed version of NP developed by the end of Chapter V gives us a way of understanding self-ascription in terms of self-identification, one that does not appeal to IEM or accept Shoemaker's regress argument in its favor. However, the view does have the rather interesting--and to some no doubt unpalatable--theoretical consequence that first-person misidentification always remains a possibility. That is, one could find oneself in the unfortunate position of denying that a token of the first-person pronoun or its mental analogue refers to oneself, even if that token participates in an introspectively conscious self-ascription. Arguably clinical data regarding schizophrenics who suffer from thought insertion suggest that this does in fact happen. In the end, however, this consequence illustrates an important virtue of NP: The persistent possibility of self-ascriptive error in both accidental and nonaccidental instances shows that they differ only in degree and not in kind with regard to self-identification. As a result, we

have a much better chance of providing a unified account of self-ascription that does not rest upon a mysterious but ineliminable assumption about the privileged access one has to one's self.

Chapter II. The Problem of Self-Identification

"I must disappear in order that the search for myself be successful."

--Stanley Cavell, The Claim of Reason¹

§1. Introduction

In *The Concept of Mind*, Gilbert Ryle vividly describes what he calls the "'systematic elusiveness' of the concept of 'I'":

Pronouns like 'you', 'she', and 'we' feel quite unmystifying, while 'I' feels mystifying. And it feels mystifying, anyhow in part, because the more the child tries to put his finger on what 'I' stands for, the less does he succeed in doing so. He can catch only its coat-tails; it itself is always and obdurately a pace ahead of its coat-tails. Like the shadow of one's own head, it will not wait to be jumped on. And yet it is never very far ahead; indeed, sometimes it seems not to be ahead of the pursuer at all. It evades capture by lodging itself inside the very muscles of the pursuer. It is too near even to be within arm's reach. (1949, p. 186)

Ryle's depiction of the mysteriousness of the first person vividly captures the difficulty of providing a satisfying philosophical account of thoughts we have about ourselves that we would normally express with the first-person pronoun. The mystery stems in large part from the relative ease with which we think about ourselves. However

¹ Found on p. 352.

difficult it may be to capture the elusive referent of 'I', the pronoun pours out of our mouths--nothing seems more natural than to ascribe properties to ourselves and to worry about our past and future exploits. The sheer number of thoughts one has about oneself on any given day must be staggering, and in most cases one is trivially aware that one is their subject. When I believe that I have lost my keys, to find them I do not first go looking for myself. When asked to explain how we so effortlessly know whom we are thinking of, however, we are at a loss as to what to say, besides, perhaps, that we just do.

Interestingly, errors do occur on occasion. I have had the experience of sitting in a small movie theater and thinking that the person whose head is casting a shadow on the screen should hunker down only to realize that it was I who was obscuring the picture all along. Together, the phenomena of first-person thoughts and instances in which errors arise raise a host of puzzling questions. How can one both be so secure in the subject of some of one's own thoughts yet fail to recognize oneself as the subject of others? How are we to account for the intuitive difference between thoughts one would express with the first-person pronoun and those that one would express with other referring expressions, and what do those abilities consist

in our amount to? What grounds our ability or abilities to get a hold of ourselves in thought? Does the self-ascription of properties and states differ in kind from the ascription of properties and states to others?

Properly addressing these questions requires great care, for the power of our intuitions regarding the first person greatly outstrips the depth of our understanding of it. To bring intuition and understanding into better balance, we must first secure a firmer grasp on the phenomena we wish to explain beyond its intuitive pull. Accordingly, in the next section I examine a set of examples of a type often associated with Heri-Nector Castañeda and widely appropriated in the literature for a host of philosophical purposes. In §3 I then turn to Roderick Chisholm's attempt to account for the Castañeda-style data in terms of a basic and unanalyzable ability to take oneself as an intentional object. Finding that solution unsatisfactory but illuminating, I devote §4 to elucidating its shortcomings and insights. In §5 I present the broad outline of a competing account that considers all self-ascriptions as susceptible to self-misidentification. I conclude this chapter with a challenge to my sketched proposal: though it provides a satisfactory explanation of a large class of thoughts one can have about oneself, it

fails to capture the way in which first-person contents intuitively appear fundamental, both in the particular functional role they play in an individual's cognitive economy and the foundational role they play in the ascription of properties to oneself and others.

Ultimately, we will see that demystifying the problem of first-person thought hinges upon understanding what it is for one to identify oneself.

§2. Framing the Problem: Characterizing First-Person Thought

One often has thoughts² about oneself. I, for example, regularly believe myself to be hungry, five blocks from home, married, reachable by phone or e-mail, 6'5" tall, thinking that a certain philosophical puzzle is thorny, and so on. I express many of these thoughts with speech acts involving the first-person pronoun. For others I appeal to various alternative self-designating expressions such as my

² By thought I mean an occurrent mental state with both a content and a mental attitude toward that content (see, e.g. Rosenthal 1993). For example, I can hope that it's raining, fear that it's raining, wonder if it's raining, etc. In each case the content--that it's raining--remains constant while the mental attitude differs. Though the ensuing discussion will focus upon belief, I intend both the problem and the subsequent approaches to the problem to extend to other propositional attitudes as well (wishing, doubting, wondering, etc.). I should also mention that, for the most part, I will speak of beliefs and thoughts interchangeably, though an important distinction should not be overlooked: thoughts are occurrent mental states whereas beliefs can be (and no doubt are largely) dispositional. This finer distinction can be overlooked for present purposes. In any event, this characterization of thought relies upon no particular theory of content.

proper name, an appropriate definite description, or a demonstrative phrase.

Those thoughts one can have about oneself that one would normally express with the first-person pronoun ostensibly seem unlike others one can have about oneself, since they resist reduction to other types of intentional state. We can bring this apparent resistance into a bit more relief by looking to a type of belief ascription often employed by Castañeda (1966, 1968, 1969, among others). Let's assume that a subject, Vivienne, believes that someone has won the New York Lottery. Let's further assume that she has long held that, as a rule, lottery winners are lucky. As it happens, Vivienne herself is in fact the lucky winner.

How are we to ascribe beliefs to her in this case? It seems that we could plausibly do so in primarily three ways:

- (1) The winner of the New York Lottery believes that (whoever is) the winner of the New York Lottery is lucky.
- (2) The winner of the New York Lottery believes of some person who is the winner of the New York Lottery that she is lucky.³

³ Chisholm, whose 1981 discussion I follow in elaborating the current example in the text, casts his de re formulation (our [2]) in explicitly quantificational terms. He would write (2) as: "There is

- (3) The winner of the New York Lottery believes that she, herself [she*⁴] is lucky.

Grammatical awkwardness aside, (1) is traditionally referred to as a case of belief de dicto--a belief where the believing relation holds between a subject and a sentence or proposition.⁵ (2) is an instance of belief de re. Here, the subject has a particular object or individual in mind of whom she wishes to ascribe luckiness. Example (3) represents a case of what is often referred to as a belief de se. That is, (3) ascribes to Vivienne the belief that she is the winner of the lottery in a way that

an x such that x is identical with the winner of the New York Lottery and x is believed by x to be lucky." My example as worded in the text does not differ in import from Chisholm's, I believe, and more closely approximates actual folk psychological belief ascriptions.

⁴ Pronouns with asterisks ('he*', 'she*') as they appear in the self-consciousness literature function as notational devices originally devised by Castañeda in a series of papers (1966, 1968, and 1969 among others). The asterisk is intended to indicate the presence of an indirect reflexive or what he calls the "quasi-indicator" use of the pronoun (e.g., 'she*' = 'she, herself'), since English lacks any such distinct form. However, to avoid the appearance of allegiance, I will resist this popular notation in favor of the more pedestrian 'she, herself' form.

⁵ Quine (1966) would term this form of belief "notional" and purportedly de re beliefs "relational". Cf. also Chisholm (1981). Sentences and propositions need not be the only belief relata; some prefer statements or sentence tokens in the head (Devitt 1984) or out of it (Quine 1981). These distinctions need not worry us here. Since nothing hinges on the characterization of propositional attitude contents (at least at the moment), those who don't want to place much stock in the de dicto/de re distinction in light of the concerns voiced by Quine (1966, 1981), Dennett (1987), and Devitt (1984), among others, need not bristle at my examples. No great fan of the distinction myself, I employ it only to give contour to the unique character of a certain class of self-ascriptions and nothing much will hang on them in my overall argument.

makes perspicuous her recognition that she believes something about herself.

This last ascription seems to differ from the other two in an important respect. Neither (1) nor (2) entails (3), for (3) can be false while either (1) and (2) are true. This is fairly obvious with the de dicto reading since, ex hypothesi, in this case Vivienne's belief "accidentally" ends up being about her by being about all lottery winners, for she believes that lottery winners in general are lucky regardless of any current beliefs she may have about her own status as a winner. Indeed, the parenthetical 'whoever' indicates that she has no particular person in mind to whom she ascribes the property of being lucky.⁶ The de re reading allows for a similar gap in recognition. In (2), Vivienne could fail to recognize that her thought is about herself even in thinking of some particular person as the winner: She could see a news flash on television, and, without

⁶ Cf. Donnellan's (1966) distinction between what he calls "attributive" and "referential" types of reference. One could argue that the case as I have constructed it is unclear as to whether Vivienne thinks that she, herself is the lottery winner. After all, perhaps she is the kind of person who takes every opportunity to inform others about her good fortune. That is, she may express her request for subway tokens as "The winner of the New York Lottery would like fifteen dollars on her Metrocard, please." I admit the ambiguity--indeed welcome it--since the key point is that the case can be ambiguous, whereas it appears that the de se ascriptions do not permit the possibility of her failing to think that she, herself is the winner.

realizing that it is she who is shown walking to the bank, Vivienne could think of that person that she is lucky.⁷ In (3), however, the presence of the indirect reflexive 'she herself' is intended to make explicit Vivienne's recognition that she believes something about herself, hence precluding the failure of self-recognition possible in both (1) and (2).

We can put the point another way. Since the indirect reflexive 'she, herself' in (3) is a pronominal phrase, we can ask how a given token of it relates to its antecedent. It does, of course, depend upon its antecedent for its designation, for the antecedent signals to whom the belief is ascribed. However, merely substituting the expression 'winner of the New York Lottery' for 'she, herself' in (3) results in:

- (4) The winner of the New York Lottery believes that the winner of the New York Lottery is lucky;

which amounts to (1), or the de dicto reading, and hence reopens the possibility of mistakes of self-recognition. Substituting other co-designating terms seems to fare no better, for replacing 'she, herself' with 'Vivienne', for example, results in:

⁷ Perry terms this possibility a failure of "documentary self-recognition" (2000, p. 333). Such examples are legion. Cf. Perry's often-quoted "shabby pedagogue" example from Mach (1914), and Lewis's (1983) use of Kaplan's (1989) example of one's pants being on fire.

- (5) The winner of the New York Lottery believes that Vivienne is lucky.

But here again, we face the possibility that Vivienne will fail to recognize that she is thinking about herself, for it seems plausible that she could affirm that Vivienne is the lottery winner while denying that she, herself is Vivienne. (We need only imagine, for example, that her denial results from memory loss following a particularly sharp blow to the head.) Moreover, substituting a demonstrative--even if reduced to a pointing or referring gesture--merely converts (3) to an instance of (2) or belief de re, again reintroducing the possibility that she will think that the winner is someone other than herself.⁸

What does it mean to say that (1) and (2) allow for the possibility that Vivienne could fail to recognize that her thought is about her whereas (3) does not? Describing Vivienne's potential for error in (1) and (2) as one of misrecognition misdescribes the case a bit. Recognition involves, at least in part, re-cognition--namely, a previous familiarity with the object recognized.

'Misidentification' is the more appropriate term, for to

⁸ The failure of substitutivity in cases of beliefs about oneself, including the appeal to the 'he, himself' locution as the distinguishing mark of these particular types of examples, was first noticed by Geach (1957/1972). As Geach correctly points out, substituting coreferential terms outside of the scope of believes--'Vivienne', say, for 'the winner of the New York Lottery' in (3)--will not affect its truth-value.

identify something is to specify or to categorize it regardless of its familiarity. In both (1) and (2) Vivienne could assume that the winner is entirely unknown to her.

What, then, does misidentifying oneself amount to? In tokening each thought ascribed to her, Vivienne succeeds in referring to herself, and she therefore has a thought about herself. However, she could deny that she refers to herself in the case of either (1) or (2). That is, it seems possible to ascribe to Vivienne the denial of (3) as well:

- (6) The winner of the New York Lottery does not believe that she, herself is lucky.⁹

This possibility of ascribing (1) or (2) and (6) will receive a much more detailed treatment in Chapter III, but for the moment we can glean from these examples that a subject can misidentify herself if for a given thought that she would normally express with some predicate F and some self-designating name, definite description, or demonstrative phrase a, a subject can assent to "a is F" while denying that she, herself is F.¹⁰

⁹ Alternatively, "The winner of the New York Lottery does not believe that she, herself, is the winner of the New York Lottery."

¹⁰ I frame my account of misidentification and the ensuing discussion in terms of singular self-ascriptions to keep it as clear

Basically, we want to capture the key feature of cases like (1) and (2)--namely that one can have a thought about oneself yet deny that that is the case. With de se belief, such a denial has no intuitive plausibility, for this amounts to ascribing to a subject the belief that she, herself is F along with the denial that she, herself is F. Such a subject would obviously harbor explicitly contradictory beliefs, bringing either her grasp of English or the soundness of her mind into question.

However intuitively implausible a competent subject's denial of a first-person self-ascription may seem, ruling out the possibility of the fairly common phenomenon of self-misidentification illustrated in (1) and (2) risks denying data useful for understanding how we think of and refer to ourselves. Accordingly, to mark this important distinction I will call thoughts and expressions of thoughts exemplified in (3) instances of nonaccidental self-ascription and those in (1) and (3) instances of potentially accidental self-ascription--or simply accidental self-ascription for the sake of brevity.¹¹ Cases in which one accidentally self-ascribes a property

and simple as possible, but it is intended to generalize to more complex ascriptions involving multiple names and many-place predicates.

¹¹ My terminology is inspired by but not entirely faithful to Bermúdez (1998).

seem to allow for a subject to misidentify herself; nonaccidental cases do not. The problem of first-person thought amounts to explaining this intuitive difference.

§3. Chisholm: The Irreducibility Argument and Fundamental Intentionality

In the last section we characterized the accidental/nonaccidental self-ascription distinction in terms of the possibility of misidentification. We characterized misidentification in terms of a subject's denial that she, herself possessed a property that she in fact ascribes to herself. So characterized, we further find a strong intuitive pull against this possibility in nonaccidental cases. Characterizing the distinction in this way, however, relies upon an understanding of a subject's believing that she, herself is F.

What is it for a subject to believe that she, herself is F? Roderick Chisholm (1981) finds belief ascriptions including the indirect reflexive quite puzzling. He is most taken by the irreducibility of ascriptions such as (3) to those exemplified by (2)--the fact that de se belief cannot be explained in terms of de re belief.¹² As we saw

¹² The irreducibility conclusion, though popular, is not uncontroversial. Boër and Lycan (1980), for one (or rather two), disagree. They argue that de se attitudes can be reduced to attitudes de re (or "attitudes de their owners," p. 432). We should note,

above, (2) does not imply (3) because (2) can be true while (3) false, as demonstrated by instances of misidentification. Chisholm believes that this "fact of intentionality" provides the raw materials for a solution to the problem of intentionality in general, or the problem of what makes a belief about a particular person about her.¹³

If de se beliefs cannot be reduced to belief de re, Chisholm argues, they must therefore constitute a separate category of intentional state. Moreover, the puzzles arising from such nonaccidental states result from ignoring their special status in favor of attempts to reduce them to accidental ones. Accordingly, Chisholm suggests that we abandon that strategy in favor of its converse: We should view cases of accidental self-ascription as special instances of nonaccidental self-ascription.

Taking nonaccidental self-ascriptions as basic involves viewing belief as a relation between a believer and a property that is ascribed or attributed to oneself. One can attribute properties in two ways, either directly

however, that they explicitly and unabashedly assume the plausibility of de re belief in general. In any event, Boër and Lycan do offer a brief but useful summary of the various uses to which this irreducibility claim has been put (pp. 429-430).

¹³ Chisholm credits this question asked in this way (viz., with its particular emphasis) to Wittgenstein (1970). Framing the question of how a belief is about an individual in terms of how a word refers merely passes on the problem.

or indirectly. Chisholm defines direct attribution in terms of the 'she, herself' locution: For a subject to directly attribute the property of being F she must believe that she, herself is F (1981, p. 28). Indirect attribution occurs in cases where one ascribes a property F to an object other than oneself. To do so, one directly ascribes a property that entails that some object is F, a property that uniquely specifies that object by way of some particular identifying relation the subject bears to it.

A slight variation on our previous example can perhaps lend additional clarity. Imagine that Bart, a friend of Vivienne's, has instead won the New York Lottery and that Vivienne believes that this is the case. What makes Vivienne's belief that Bart is lucky about Bart, under this view, is (a) her standing in a unique, identifying relation with him (e.g., the relation of being the person who is currently sitting directly to her right on the bus); and (b) her indirectly attributing to him the property of being lucky. The attribution in (b) is indirect because she directly attributes to herself the property of sitting to the left of one and only one person and a person who is lucky. Hence, all thoughts attributing properties to others can be reduced to thoughts attributing properties to oneself, making the believing subject "the

primary object of all belief, and analogously for the other intentional attitudes" (Chisholm, 1981, p. 37).

Given this picture, Chisholm's explanation of the difference between accidental and nonaccidental thoughts about oneself, and of the resulting philosophical puzzlement, is quite straightforward. In nonaccidental self-ascriptions one directly attributes a property to oneself; in accidental cases--cases where one is not aware that one is thinking of oneself--one actually indirectly attributes a property to oneself without directly attributing it (p. 34). In (2), misidentification occurs because Vivienne indirectly and accidentally attributes the property of luckiness to herself by self-ascribing a property under a specifiable relation that in fact picks her out, such as the property of watching one and only one person on the news and one who is lucky. And in imagining possible scenarios of self-misidentification, we are really considering indirect ascriptions involving various identificatory relations that one would normally bear to objects other than oneself.

§4. Revisiting Chisholm's Conclusion

Chisholm's strategy for reaching his conclusion is twofold. First, he appeals to intuitions elicited by common examples

of belief ascriptions to isolate a particular type of intentional state as unique. Having limned what he takes to be the appropriate data, he then develops an account of how one takes oneself as an intentional object--and ultimately how one shares an intentional relation with any object at all--based upon it. Doing so provides him a way of explaining the distinction between accidental and nonaccidental self-ascriptions, including the commonsense errors of self-misidentification to which we are often prone. Let's examine each move more closely, taking the second one first.

Chisholm argues that the irreducibility of belief ascriptions involving the 'she, herself' locution demonstrates that one's ability to take oneself as an intentional object is fundamental and not further analyzable in terms of any relation one could be said to have with oneself. Broadly speaking, he wants to capture the notion that since I think and speak about all things relative to me, thinking and speaking about myself must be taken for granted.¹⁴ All beliefs about oneself and others

¹⁴ Lewis (1983) puts forward a quite similar view, differing mainly in the details and particularly his reliance upon possible worlds. For Lewis, de se beliefs are fundamental because they locate oneself within a set of possible worlds and then further spatially and temporally within that particular world. Ascribing a property to some object or individual involves self-ascribing a property under a world-locating description that either captures it's "essence" or the

therefore reduce to direct ascriptions of properties to oneself.¹⁵

But if direct attributions of properties must be taken as given, how are we to understand them? Don't we still need to explain how one's attributions of properties directly to oneself succeed in being about oneself? Identification is the complement of misidentification, so perhaps the possibility of self-misidentification indicates that self-identification succeeds in instances of direct attribution of properties. Don't we therefore need an account of how one identifies oneself, including one's apprehension of oneself "enabling one to pick out the self from all other things" (Chisholm 1981, p. 37)?

Chisholm considers this objection mistaken. He believes that it

confuses direct and indirect attribution. For it presupposes that, in order to attribute a property to myself directly, I must be aware of the kind of identifying relation I bear to the subject of attribute when I attribute something to it indirectly. (p. 37)

In Chisholm's view, then, properly speaking there is no direct self-identification at all, for object

subject's relation of acquaintance with it (p. 155). Fundamentally, the relation of acquaintance one has with oneself is one of identity, what Lewis calls "a relation of acquaintance par excellence" (p. 156).

¹⁵ Jaegwon Kim has characterized this type of view as referential or intentional solipsism.

identification is the product of a specifying relation by which one indirectly attributes a property. Requiring that one identify oneself to attribute a property directly would demand that one be able to specify a relation that picks oneself out uniquely, but this conflates the requirements for direct and indirect attribution. "For in directly attributing a property to oneself," he writes, "one need not thereby single out an identifying property of oneself" (p. 45).

What motivation have we for endorsing Chisholm's strategy of taking the ability to take oneself as an intentional object as theoretically primitive? He asks and answers this question himself:

But to the question: "What makes his direct attribution of a property to himself an attribution of a property to him?" there can be no answer at all, beyond that of "He just does--and that is the end of the matter!" Do we have here, then, a difficulty that is unique to the present theory? It is important to see that every theory of reference and intentionality is such that, at some point, it must provide a similar answer: "It just does." (p. 32)

This answer, however, and with it Chisholm's proposal, hardly satisfies. After all, we seek an explanation of how one thinks of oneself directly, but this is the question to

which Chisholm rejoins "One just does!"¹⁶ We may ultimately arrive at a similar conclusion, but we should nevertheless exhaust the theoretical options that attempt to provide an explanation of this ability. Otherwise, we resign ourselves to the position of Ryle's child and to the mysteriousness of the first-person.

Moreover, though Chisholm explains the philosophical puzzlement surrounding nonaccidental self-ascriptions in terms of cases in which one indirectly ascribes a property to oneself, he provides very little in the way of an explanation as to why one indirectly self-ascribes a property. For if one possesses the fundamental ability to attribute properties directly, how is self-misidentification possible at all? In taking nonaccidental cases of self-ascription as fundamental, Chisholm needs to provide a better account of the accidental cases that highlight the intensional behavior of the intentional, those that he describes as providing a "kind of criterion of the psychological" (p. 25). In the end he offers at best a description of accidental and nonaccidental cases instead of the explanation we seek.

¹⁶ I should also note that Chisholm likewise assumes one's ability to grasp or conceive properties, propositions, and states of affairs (p. 28). For the sake of argument I will not take issue with this second assumption.

A further objection looms. When I believe that I am lucky, I directly attribute the property of luckiness to myself. In doing so, do I thereby believe something true? The answer to that question ostensibly depends upon the properties I possess. But if the content of the attribution is the property of luckiness and the object is myself, we seem to have no proposition-like object that has truth conditions, and hence the belief can be neither true nor false. Anticipating this objection, Chisholm argues that we consider direct attributions that one would express with the sentence "I am F" as "used with truth in English if and only if its utterer is F" (p. 44, emphasis his).

This is misleading, however. Though Chisholm distinguishes between using an expression with truth and using it with sincerity (since presumably one often believes false things sincerely), truth is not a matter of the way in which expressions are used. The truth-value of "I am F" does not depend upon how one uses the expression, for presumably one uses it to assert something about oneself in the same way, if done so sincerely, whether one actually is F or not.¹⁷ Without a clearer sense of what Chisholm means by using an expression with truth or

¹⁷ To put the point another way, no speech act has the particular illocutionary force of sincerely asserting a falsehood.

falsity, we have no grasp on how to understand folk-psychological explanations of belief in which the notion of believing truthfully or falsely plays a significant role.¹⁸

As an added difficulty, Chisholm's view generates quite complicated explanations of rather commonsense exchanges of information between speakers. Suppose Vivienne, unsure about her luck, asks Bart, "Did I win the lottery?" Bart responds, "Yes, you are the winner." According to Chisholm, Bart must interpret Vivienne as wondering which property she should ascribe to herself, and he must infer from the property she does self-ascribe that she asks this of him. To answer, he must indirectly ascribe to Vivienne the property of being the lottery winner, which means he must directly ascribe to himself a property that singles her out for attribution, the property of addressing one and only one person sitting next to him on the bus, for example (pp. 48-49). Each party must, therefore, infer from what the other says of herself or

¹⁸ This problem plagues Lewis's account as well (Lewis 1983). For him, speech acts of the form "I am F" can represent a subject's locating herself both in logical space, or a set of possible worlds, and in space and time within a particular world. Possible worlds consist of sets of true propositions, but according to Lewis knowing which propositions are true doesn't suffice to locate oneself within a world. But if "I am F" were true, then presumably it would serve as a world-delimiting proposition on par with any other. Lewis also claims that one can have nonpropositional knowledge, but since presumably knowledge crucially involves truth, it's not clear what nonpropositional knowledge could be.

himself to her or his own case.¹⁹ Intuitively, it seems that Vivienne asks something directly of Bart and that Bart speaks only of her. Though it may sound quite odd to claim that one always speaks and thinks about oneself, whether answering one's own questions or the questions of others, philosophical accounts of the apparently commonsensical often belie the simplicity of appearances. This point is not so much an objection as a worry, a consequence to be avoided if at all possible.

Perhaps Chisholm's conclusion and its resulting mysteriousness can be resisted by reconsidering his account of the data of self-ascriptions. The irreducibility of beliefs ascribed with the 'she, herself' locution to other sorts motivates him to assume that a subject has a fundamental ability to take herself as an intentional object. Should nonaccidental self-ascriptions fail to prove as distinctive as Chisholm suggests, we can avoid being forced into his assumption.

Recall that the intuitive pull of regarding nonaccidental self-ascriptions as different in kind from their accidental counterparts received additional support from the entailment relations between (1), (2), and (3).

¹⁹ Stalnaker (1981) makes a similar objection to the proposals given by Lewis and Perry (1977, 1979).

Those relations could also be elucidated in terms of the ascriptions resulting from substituting co-designating terms for 'she, herself' within nonaccidental ascriptions. Doing so converts nonaccidental ascriptions into accidental ones by opening up the possibility of self-misidentification. Hence nonaccidental self-ascriptions are different in kind from their accidental counterparts.

But failure of substitutivity in propositional attitude contexts presents no particularly new problem or a problem unique to nonaccidental self-ascriptions. Frege famously introduces the notion of sense to explain how one could have distinct and seemingly contradictory beliefs about the same referent. One could harbor beliefs about both the Morning Star and the Evening Star while lacking the belief that the Morning Star is the Evening Star. Accordingly, given the truth of:

- (7) Vivienne believes that the planet Venus is the Morning Star.

we can substitute the co-designating term 'Evening Star' for 'Morning Star', which results in

- (8) Vivienne believes that the planet Venus is the Evening Star.

Though Venus is the Evening Star as well as the Morning star, Vivienne might well deny Venus' evening visibility,

and hence (8) could be false while (7) is true. As with (1) and (2), substituting any co-designating term within the belief context can alter the truth value of the ascription, and this result squares with our expectations given the ample evidence of the general opacity of propositional attitudes.

Perhaps we can further distinguish the peculiar nature of nonaccidental self-ascriptions beyond their susceptibility to Frege-style puzzles by moving from oratio obliqua belief ascriptions couched in opaque propositional attitude contexts to oratio recta belief expressions outside such contexts. The pronominal phrase 'she, herself' functions grammatically as an indirect reflexive pronoun: It is indirect since it is affected by the action of the propositional attitude verb (believes, wants, hopes, etc.) that precedes it, and it is reflexive because it includes the reflexive form of the pronoun (ending in '-self' or '-selves') whose grammatical job is to refer back to the subject.²⁰ The direct reflexive, oratio recta

²⁰ Reflexive pronouns in English traditionally serve as the direct object of a reflexive verb. Hence, English, unlike, e.g. Greek, has no particular indirect reflexive form. The 'she, herself' formulation is an exotic grammatical creature, for it consists of both the subject and reflexive (which is an object) form of the third-person pronoun. Castañeda (1967) is often credited as the first to notice it, calling it (or his notational variant of it) a "quasi-indicator." He believes that it serves to refer to the way in which a particular subject thinks of herself. Perry (2000) characterizes Castañeda's

equivalent is the first-person pronoun. With (3) we ascribe a belief to Vivienne that she would express as:

(9) I am lucky.

The temptation is to construe self-ascriptions like (9) as expressing a unique way Vivienne thinks of herself, a way in which no other can think of her.²¹ Ascriptions (1) and (2), in contrast, involve representations of Vivienne that we might describe as available to others. (1) and (2) ascribe beliefs that Vivienne might express as:

(10) V is lucky.²²

where 'V' refers to her. Bart can truthfully believe (10) as well, but were he to believe (9) he could believe something false of himself.

Moving from indirect to direct discourse still doesn't succeed in singling out the class of first-person

account of the role of quasi-indicators as senses that refer to senses--that is, a third-person term with a sense that refers to a sense that can only be grasped in a first-person way. He further describes Castañeda's goal as attempting to make a distinction between sense and meaning. I am unsure what to make of or how to understand this latter explanation. See also Anscombe (1975).

²¹ Castañeda, for one, draws this conclusion, arguing that "more generally, no first-person proposition is identical with any third-person proposition, if both are in oratio recta" (1966/1994, p. 163, emphasis in original).

²² By condensing (1) and (2) into (7) we obscure the distinction between de dicto and de re belief; still, we do preserve the property they share that initially interested us--namely, that they both allowed room for Vivienne to misidentify who her beliefs were about since, intuitively, she might deny that she, herself is lucky by denying that she, herself is V.

thoughts as unique, however. Here again, we run afoul of substitutivity. Though (9) when thought or said by Vivienne and (10) have the same truth-conditions, intuitively Vivienne could believe either one while denying the other. Even if Vivienne does believe that she, herself is the lottery winner, if she does not believe that she is V, then she won't believe that V is lucky. Similarly, though Vivienne may believe that V is the lucky winner, she may not believe that she, herself is V and thus may not token the belief that she would express with (9). This is no different from Vivienne's believing what she would express with

(11) The planet Venus is the Morning Star.

while denying

(12) The planet Venus is the Evening Star.

though they share truth conditions.

Nevertheless, looking at a subject's beliefs in terms of their direct discourse expressions does seem to offer a way to distinguish nonaccidental self-ascriptions from others. The belief that Vivienne would express with (10) exemplifies a range of cases where misidentification is still possible--that is, where, intuitively, a subject can have the belief yet fail to recognize that she is its

subject. This indicates that tokening a belief that one would express with a self-designating expression such as a proper name or definite description isn't sufficient for one to identify oneself as the subject of the description.²³ Tokening a self-ascription that one would express with the first-person pronoun does seem sufficient for self-identification. No room appears to remain for one to deny that one thereby ascribes a property to oneself.

But what lies behind this appearance? After all, we do often seem to succeed in identifying ourselves as the subject of potentially accidental self-ascriptions. Perhaps that success has much in common with nonaccidental instances. Does thinking about oneself "in a first-person way" in fact differ from other ways in which one thinks about oneself, and does the impossibility of self-misidentification imply that self-ascriptions made with the first-person pronoun or its mental analogue involve no self-identification whatsoever?

²³ In fact, often as the description becomes more precise, it appears that the chance of misidentification increases. Vivienne could be a geneticist, routinely handling several detailed DNA maps. It clearly seems possible that she could believe on the basis of a particular DNA map that a certain person has committed a crime while not believing that she, herself is a criminal, even though the DNA map is her own.

§5. An Alternative: The Naïve Proposal of Self-Identification Introduced

In the last section we saw that Chisholm argues from the entailment relations between nonaccidental self-ascriptions like (3) and accidental ones like (1) and (2), to the irreducibility of de se belief to either belief de re or de dicto. Irreducibility warrants taking de se belief as basic and defining all other forms in terms of it. But Chisholm teaches us, perhaps inadvertently as far as he is concerned, that the interesting cases of self-thoughts are those in which misidentification of oneself is possible, the ones in which one has a thought that predicates a property of oneself but one could readily deny that one has it.

Let's focus, then, on just singular accidental self-ascriptions in which misidentification does occur. At the risk of naïveté, I recommend endorsing--at least initially--a fairly simple proposal: Accidental ascriptions occur when one succeeds in referring to oneself but fails to identify oneself as the referent. The belief that Vivienne would express with

(10) V is lucky.

is just such a case of self-reference without self-identification, for the accident my terminology is intended

to capture is her accidental reference to herself. In contrast, Vivienne's use of 'I' to self-ascribe the property of being lucky as in

(9) I am lucky.

indicates that she intends to refer to herself. Since she has succeeded in "getting a hold of herself in thought," as it were, referential accidents don't seem to present a worry.

But what does it mean for one to intend to refer to oneself or to "get a hold of oneself in thought"? Ascribing a property to something requires picking that thing out in some way. In the case of self-ascriptions, the object to be picked out is, of course, oneself. Accordingly, nonaccidental cases of self-ascription are those in which one has succeeded in picking oneself out, and accidental cases involving self-misidentification are those in which this picking out has gone awry in some way. To be a little more specific, as we saw above the thought expressed by (10) could escape Vivienne's recognition as applicable to her because she fails to recognize 'V' as specifying her--that is, she fails to believe that she can be referred to with the expression 'V', either by herself or by others.

In looking for the kind of information capable of grounding such self-identification, a fairly uncontroversial source stands ready to hand. Like each of us, I currently possess many commonsense identificatory beliefs about myself, including my name, the names of my parents, my birth date, my height, weight, and eye color, my Social Security number, my current address, my tendencies of mood and habits of mind, my political convictions, the state of my bodily needs, and so on. Roughly speaking, such beliefs are those I would offer when considering the question "Who am I?". Perhaps none of these beliefs alone allows me to pick myself out, but in conjunction (or disjunction) they can serve as a means by which I can "get a hold of myself in thought" with a high degree of precision. Ascriptive accidents occur when a subject's battery lacks certain beliefs or contains false ones.²⁴

Picking out an object or individual, at least in the identificatory sense I'm using it here, should not be confused with referring to it. As Kripke (1980) teaches us, a speaker or thinker can refer to something even if she

²⁴ The view I begin with here and fill out in more detail later shares many similarities with the view put forward by Rosenthal (2003) and Dennett (1991). We shall give these views the attention they deserve in Chapter V.

possesses little or no specifying information about it.²⁵ One's beliefs about oneself, in contrast, taken together do identify oneself. And one's self-specifying beliefs have implications for the inferences one makes and the actions that one performs. For Vivienne to identify herself as the winner of the New York lottery simply is for her to be disposed to token particular thoughts and to act in ways appropriate to lottery-winners, including checking and re-checking the numbers on her ticket, cashing it in, telling her acquaintances about her good fortune, hiring an accountant, etc. Had she not believed that she was the winner of the lottery--had she not identified herself in that way--the reward due her would have gone unclaimed by her. Descriptive beliefs constitute a self-identifying battery just in case they interact with other intentional states, particularly desiderative ones, to produce self-specifying behavior.

The naïve proposal--hereafter NP--explains nonaccidental self-ascriptions in the same way. One's first-person self-ascriptions also draw upon one's battery of descriptive information to identify their subject.

²⁵ Kripke (1980) makes the point in a compelling way with his famous Gödel/Schmidt example (pp. 83-91). The upshot is that when one erroneously associates a description with the name of an individual other than the one who satisfies it intuitively one still refers to the bearer of the name.

Indeed the truth of (9) depends upon Vivienne herself tokening it, whereas anyone can truthfully token (10). But this disparity follows from the grammatical role of 'I'-- one cannot help but refer to oneself in tokening thoughts or speech acts involving the first-person pronoun or its mental analogue. We have seen, however, that self-reference is dissociable from self-identification. And the disparities between (9) and (10) that we may want to explain in terms of asymmetries in first- and third-person access can instead be attributed to the richness of descriptions appealed to. With 'V' Vivienne has but one description in play ("the bearer of the initial 'V'" or perhaps "the person called 'V'"). With 'I', however, she can draw upon her entire battery, including the rule that the English word 'I' refers to the thinker or speaker of the thought or speech act in which it or its mental analogue appears. One's battery of descriptive information about oneself will always be richer than that possessed by another, for, as it were, one spends significantly more time with oneself than anyone else. Moreover, just as no two subjects have identical histories, no two subjects have identical batteries. This does not mean, however, that one is automatically familiar with all the ways in which one

can pick oneself out--indeed, one is no doubt acquainted with very few of them.

Like accidental self-ascriptions, nonaccidental thoughts about oneself expressed with the first-person pronoun in effect reveal the description or descriptions one currently believes to be true of oneself. 'I' (in Vivienne's mouth) and 'V' each refer to the same thing, but they provide different ways of identifying that referent. Hence, on this simple proposal, both successful accidental and nonaccidental self-ascriptions are possible in virtue of one's being able to specify the referent of a given ascription token, whether they involve the pronoun 'I' or some other self-designating expression. Specifying the referent amounts to being able to pick it out, and one picks oneself out via a battery of descriptions that serve to identify oneself uniquely.

Accordingly, the puzzling apparent distinction between accidental and nonaccidental self-ascriptions indeed proves to be only apparent. One applies one's ability to identify a referent in both cases, with the difference a matter of descriptive resources that can be brought to bear in each case. This is to say that nonaccidental and accidental self-ascriptions--and hence

first- and third-person intentional states--differ only in degree and not in kind.

Ryle (1949) puts forward a quite similar view of self-ascription to dispel the myth of privileged access, or the view that "minds possess powers of apprehending their own states and operations superior to those they possess of apprehending facts of the external world" (p. 154). He argues that we come to know and to believe things about ourselves in much the same way that we come to know and to believe things about others. Self-knowledge results not from some powerful "inner awareness" which makes the mind transparent to itself but instead from considering ourselves in retrospect, by "catching" ourselves engaged in moods, thoughts, or behaviors and through mulling over past events in our autobiographies. Though it may seem that one can't catch another cursing, say, in the same way that one can catch another doing so, Ryle contends that

the differences are differences of degree, not of kind. The superiority of the speaker's knowledge of what he is doing over that of the listener does not indicate that he has Privileged Access to facts of a type inevitably inaccessible to the listener, but only that he is in a very good position to know what the listener is often in a very poor position to know. The turns taken by a man's conversation do not startle or perplex his wife as much as they had surprised and puzzled his fiancée, nor do close colleagues have to explain themselves to each other as

much as they have to explain themselves to
their new pupils. (p. 179)

However, persuasive Ryle's arguments may be, the privilege he wishes to deny the first person differs from the one we have been considering. For even if we were to agree with Ryle that one can have false beliefs about oneself, we still have the intuition that one can't be wrong about who thinks such-and-such or instantiates such-and-such properties when one nonaccidentally self-ascribes those thoughts and properties. First-person self-ascriptions might still enjoy a type of privileged identification. To put it into Ryle's own terms, a man's wife might indeed be less perplexed by her husband's turn of phrase than his fiancée, but it sounds strange to say that she's more sure now than before whether it is she, herself who awaits her turn to continue the conversation.

Nevertheless, NP honors the spirit of Ryle's proposal and thus inherits its desirable features. First, since it demystifies first-person self-reference by dissolving any hard and fast boundary between accidental and nonaccidental self-ascriptions--and, hence, ascriptions to other selves and nonselves alike--it does not have the metaphysical implications haunting views that take nonaccidental ascriptions to be fundamental. As a result, it could

easily prove commensurate with other physicalist commitments and square nicely with a complete description of the physical world in objective, third-person terms. As a further advantage, it appears more resistant to nagging, traditional skeptical problems such as the problem of other minds. For if self-ascriptions were fundamentally like other-ascriptions, we would be more resistant to granting that the first person stands in an epistemically privileged position from which to self-ascribe states and properties. In Ryle's words, in denying a difference in kind between the first and third person "we lose the bitters with the sweets of Solipsism" (1949, p. 156), no longer forced to regard each individual mind as a "metaphysical enclosure."²⁶

Perhaps most importantly, this proposal provides a unified account of thoughts about oneself that can easily harmonize with accounts of thoughts in general. In particular it assigns no special status to nonaccidental self-ascriptions. This makes good theoretical sense, since any physicalist account of the mental lives of thinkers will need to explain how mental states interact causally with environmental stimuli and other states to produce and to predict behavior. Unified accounts do not face the

²⁶ I borrow this last wonderful phrase from Cavell (1976).

added difficulty of explaining the interaction of states with fundamentally different properties.²⁷

§6. Objections to NP

Whatever the benefits of NP, it faces several troubling objections. First, it must confront what Qassim Cassam (1997) calls the Misconception Problem (p. 127).²⁸ I may indeed possess a certain amount of descriptive information that I take to be self-specifying, but that information may in fact not specify me uniquely or at all. No doubt we all harbor at least a few false beliefs, and it would be the height of hubris for one to hold that none of them contributes to one's personal self-conception. I could, for example, believe that I am the son of Bertrand Russell,

²⁷ Lewis (1983) motivates his own account of properties as the uniform objects of propositional attitudes largely because such uniformity facilitates a description of the causal interactions of the attitudes (p. 134). Strictly speaking, both Chisholm and Lewis do present unified views--all thoughts are ultimately self-thoughts.

²⁸ See his 1997, p. 127. There he offers a similar objection to views that take the fundamental nature of self-consciousness to consist of a conception of oneself as located, shaped, and solid. As Cassam presents it, one can succeed in thinking first-personally even though one believes that one is fundamentally a Cartesian immaterial substance, despite the fact that one is actually a physical object in the world among others. In other words, one could harbor radically mistaken beliefs about what sort of thing one is yet still be able to successfully know which thing she is. Cassam takes this problem as a strong motivation for the view that first-person thought is made possible by an intuitive awareness that one is a physical object among other physical objects.

This objection also shares more than a passing resemblance to one given by Kripke (1980) against descriptive theories of names such as Russell's and Frege's. If reference can succeed despite one's associating incorrect or anemic descriptions with it, names must not then refer in virtue of their descriptive content.

whereas in fact I have no philosophical lineage to speak of. Still, even in thinking of myself in this false way I nonaccidentally self-ascribe the property of being Russell's son, and we can imagine that my behavior bears out this identification as well. It's difficult to see how I can pick myself out with non-self-specifying tools.

A second, similar objection arises. Imagine that Vivienne has a doppelgänger inhabiting a remote alternative world exactly like her own. Twin Vivienne, or "Twivienne", identifies herself exactly as Vivienne would since she associates with herself a battery of descriptive information about herself consisting of beliefs that are identical to Vivienne's in every way. If Vivienne believes herself to have brown hair, Twivienne does as well; if Vivienne thinks of herself as an independent thinker, so does Twivienne, etc. They each therefore engage in the same inferences and actions resulting from their identificatory beliefs. How would one know--even, as it were, "from the inside"--whether one was Vivienne or Twivienne? NP seems unable to supply an answer, and hence it seems that however rich one's battery of self-identifying belief it isn't sufficient for one to pick oneself out uniquely. In contrast, agreeing with Chisholm that "each of us has a kind of privileged access to

himself" associated with the use of the first-person pronoun obviates this difficulty, since Vivienne and Twivienne will each have privileged access only to herself (1981, p. 43).²⁹

NP predicts, quite counterintuitively, that if this battery of descriptions were to be lost, one would no longer be able to think first personally. Let us imagine that following a particularly sharp blow to the head, Vivienne suffers acute memory loss. She can no longer recall her name, Social Security number, favorite ice-cream flavor, the names of her mother and father, whether she tends to vote Democrat or Republican, and so on. Under NP, the information that she has had violently removed from her memory qualifies as self-specifying information that would normally be given in response to the question that she would express as "Who am I?". It seems perfectly acceptable to suppose that she can no longer identify herself as a Republican, Vivienne, or 322-41-9989. However, intuitively it seems that she would remain able to

²⁹ This objection has much in common with Nagel's (1986) puzzle that the thought "I am TN" (in his case) is anything but trivial even given a complete objective description of all aspects of and individuals in the world from no particular point of view. Nagel worries that this "centerless" conception of the world seems to leave out "the fact that one of those persons, TN, is the locus of my consciousness, the point of view from which I observe and act on the world" (p. 56). However seductive Nagel's puzzle may initially seem, its pull quickly dissipates once one attempts to make sense of it, as Cassam nicely notes (1997, p.20).

think that she doesn't know who she, herself is or that she has forgotten her own name. Though Vivienne lacks much factual self-identifying and self-specifying information--perhaps she lacks all of it--it nevertheless seems intuitive that she would have no difficulty in distinguishing the search for her identity from that of, say, her mother's.³⁰ Hence, a battery of self-specifying beliefs does not seem necessary for one to nonaccidentally self-ascribe certain properties.

Finally, the proposal faces what we might call the Asymmetry Problem. In denying a difference in kind between accidental and nonaccidental self-ascriptions, NP seems ill-equipped to explain the intuitive difference between the bases of self-ascription. Few contemporary philosophers endorse the claim that first-person avowals are infallible, but just as few would deny that they are made on the basis of evidence. Indeed, NP attributes the apparent iron-clad security against self-misidentification that nonaccidental ascriptions enjoy as a result of one's possessing a wealth of identificatory information about oneself. But this is to say that nonaccidental self-

³⁰ Cf. Descartes' order of operations in the Second Meditation. He first deduces that he is--"I am, I exist"--and then proceeds to investigate what he is. Since the latter amounts to cataloguing one's properties, such self-knowledge seems unnecessary for at least a range of first-person tokens.

ascriptions are made on the basis of more evidence than accidental ones. This consequence runs counter to our intuitions, for as Davidson (2001) points out, we seem to rely on no evidence whatsoever when making such self-ascriptions. Asymmetry in evidence base between accidental and nonaccidental self-ascriptions may be illusory; nevertheless NP will have to explain--or perhaps explain away--this asymmetry.

Hence, the view that all of one's self-ascriptions depend in some way or other upon one's ability to pick oneself out by way of an identifying description or battery of descriptions faces great challenges. Together the Misconception Problem, the Amnesia Objection, and the Asymmetry Problem propel us toward the conclusion that we must possess some fundamental "self-ascriptive ability," an ability to think about ourselves that must precede or ground our ability to deploy descriptive information and to permit its acquisition as Chisholm believed. Resisting this tempting conclusion will require disarming these objections and addressing our deep intuitions regarding nonaccidental self-ascriptions.

§7. Conclusion

The larger question before us, then, is the relationship between ascriptions one makes with the first-person pronoun and all others. Ascribing a property F to any object or individual a amounts to tokening a thought that one would express with the sentence type "a is F," where 'a' designates the object and 'F' is a predicate that denotes some property. Self-ascriptions are those in which 'a' refers to the tokener of the thought or speech act, and nonaccidental self-ascriptions of the purportedly special sort have the form "I am F". What remains to be seen is whether nonaccidental self-ascriptions are a subset of the class of singular ascriptions more generally--including accidental self-ascriptions employing self-designating subject terms other than 'I'--or whether, as Chisholm and others contend, ascriptions of the form "a is F" fundamentally depend on nonaccidental ones.

Solving this puzzle requires investigating the degree to which self-identification is involved in both accidental and nonaccidental self-ascriptions of properties. Thus far, we have traced the two extremes of possible answers. The first is the view put forward by Chisholm that one possesses the ability to self-ascribe properties without any need for self-identification. At the opposite pole

lies NP that explains all self-ascriptions in terms of the success or failure of identifying oneself. The former seems to run afoul of accidental self-ascriptions, affording no real explanation as to why we sometimes think about ourselves yet deny that we do. In contrast, the latter excels at explaining misidentification but seems unable to account for intuitions regarding the impossibility of misidentification in nonaccidental cases.

In the chapters to follow, I apply the lessons learned at these extremes to develop an account that occupies a tenable middle ground between them. To continue the dialectical progress towards this goal, in the next chapter I take up the strongest arguments for the claim that one does not--indeed cannot need to--identify oneself in order to nonaccidentally self-ascribe properties with the first-person pronoun. The challenge to one like myself who finds such arguments unconvincing is to explain how misidentification remains a possibility, however remote, in nonaccidental cases. I contend that an account able to provide such an explanation stands the best chance of constructing a trap sufficient to draw out and to catch the mysterious and elusive 'I'.

Chapter III. Assessing Ascriptive Accidents: Guaranteed Reference and Immunity to Error Through Misidentification

"We who are self-possessed
own nothing. Even our names
get up in the night and run away."

--Richard Shelton, Journal of
Return

"I am not sure which of us it is that's writing
this page."

--Jorge Luis Borges, "Borges and
I"¹

§1. Introduction

As we saw in Chapter II, understanding the various ways in which one ascribes properties to oneself requires securing a grasp on the nature of both accidental and nonaccidental self-ascriptions. We examined two types of proposal that explain self-ascription in terms of the need for self-identification and the corresponding potential for self-misidentification that each permits in nonaccidental cases. The first, represented by Chisholm, embraces the claim that a subject possesses a basic and unanalyzable ability to

¹ The Shelton is from Selected Poems, 1969-1981, p. 1 and rather elegantly captures the sentiments of Wittgenstein, Anscombe, and Shoemaker discussed in this chapter. The Borges passage comes from Collected Fictions, p. 324. Interestingly, this final line to the story--"No sé cuál de los dos escribe esta página"--has also been translated as "I don't know which of us wrote this" (Selected Poems, p. 93). The translation quoted in the body above proves more interesting to the issues at hand since we find the writer currently amidst doubt concerning his identity, whereas the past-tense version permits the more pedestrian possibility of a faulty memory.

take herself as an intentional object. Only accidental self-ascriptions and ascriptions of properties to objects and individuals other than oneself involve what Chisholm calls indirect attribution, or predications that require a subject to be able to provide an identifying relation that uniquely specifies a particular object. In directly attributing or nonaccidentally ascribing properties a subject need not identify herself at all, and since misidentification can only result from a failure of identification, self-misidentification is in principle impossible in these cases.²

We considered a second view, which we dubbed the naïve proposal or NP. On this picture, all self-ascription involves self-identification, and self-identification depends upon one's possessing and deploying a battery of fairly commonplace self-specifying descriptive information. It follows that self-misidentification remains a possibility with all thoughts one can have about oneself,

² As discussed in the previous chapter, Lewis (1983) makes a similar assumption, though implicitly. He claims that one nonaccidentally ascribes properties to oneself under a relation of identity (what he calls a "relation of acquaintance par excellence," p. 156). Presumably, this identity would be expressed along the lines of "the thing identical with me that is F" Though relations need not be linguistic items (or even linguistically expressible) for Lewis, I do need some prior grasp of 'me' to complete the relation. Hence, just like Chisholm, Lewis thinks that since oneself anchors all attitudes one holds towards objects--oneself always occupies one position in any intentional relation to anything at all--the attitude one holds towards oneself must be taken for granted.

and hence nonaccidental and accidental self-ascriptions differ at most in degree but not in kind.

Each proposal has its strengths and weaknesses. According to Chisholm, one simply cannot misidentify oneself in nonaccidentally self-ascribing properties, and this seems to square with many of our pretheoretic intuitions regarding self-ascriptions--especially those that one would express with the first-person pronoun. But Chisholm's assumption that one simply has the ability to take oneself as an intentional object courts controversy. We have yet to understand what this assumption actually amounts to with any clarity, and we have yet to see what argumentative support can be marshaled on its behalf apart from its intuitive pull. Proponents also need to address its prima facie inability to satisfactorily explain actual errors of identification that typify accidental cases, and do so in a way that does not preclude a unified account of how mental states in general interact causally with environmental stimuli and with each other to produce, to predict, and to explain behavior.

NP faces its own set of challenges. It easily accounts for ascriptive accidents--mistakes of self-identification occur when one denies that a particular predication applies to oneself when it indeed does, and one

slips into error because of a failure to believe that one can be referred to or picked out in a particular way. And since it explains the difference between accidental and nonaccidental self-ascriptions in terms of the amount of self-specifying information brought to bear in a particular case, it offers no impediment to a unified account of property ascriptions in general. Nevertheless, it rightly earns the epithet "naïve" because in its present form it offers little insight into the connection between identification and ascription. As a result it lies open to objections concerning the type and degree of identifying information one needs about oneself to self-ascribe properties nonaccidentally, offending our pretheoretic intuitions regarding the ways in which we can err in thinking about ourselves in the first person. Any serious account of self-ascription must address the presence and force of these intuitions, even if the best theory ultimately vacates many if not all of them. In addition, the proposal's viability hinges upon its latent power to provide a much more thorough account of the relationship between self-identification in the battery sense and one's ability to self-ascribe properties generally.

The assumption regarding the impossibility of self-misidentification forms the primary ground of difference

between the two proposals. We cannot fully appreciate the strengths and weaknesses of each proposal, then, without elucidating both exactly what it means to say that one simply cannot misidentify oneself in certain cases, and why anyone would want to say it. Accordingly, I begin in §§2 and 3 with the contributions of Wittgenstein and G. E. M. Anscombe respectively to this subject. They each contend that it makes no sense to doubt the identity of the subject of certain self-ascriptions, even if it ultimately entails denying that 'I' functions as a referring expression in some if not all of its uses. Since in the end they inadequately illuminate the assumption that interests us here, I devote §4 to the subtle position of Sydney Shoemaker. Much like Chisholm, he argues that we must consider at least some self-ascriptions foundational and immune to errors of misidentification to explain how a subject could succeed or fail to identify anything at all, herself included. Though Shoemaker presents perhaps the most compelling case in favor of adopting Chisholm's position--one that seems to provide many of the theoretical desiderata outlined above while avoiding many of the associated difficulties--we see in §5 that he cannot accommodate many of the actual ascriptive errors that can befall subjects. In the end, we attain a clearer sense of

the ways in which NP must mature before it can successfully explain how we think about ourselves.

§2. Wittgenstein's 'I': First-Person Pronoun Used as Subject and Object

Wittgenstein's remarks concerning the self and the first-person, though quite influential, consist of some of his most elliptical and recondite work. Nevertheless, he does consider why one might think that nonaccidental self-ascriptions differ in kind from accidental ones. In brief, Wittgenstein maintains that whereas one can doubt who exactly is the subject of an accidental self-ascription, it is strictly nonsensical for one to call the subject of one's nonaccidental self-ascription into question. Questions about the identity of a particular subject could only arise if such self-ascriptions required a subject to identify herself in some way. Consequently, he concludes that nonaccidental self-ascriptions are distinct from all others in that the first-person pronoun used to express them does not serve to pick out or to signify a particular ascribing subject.

In the Blue Book,³ Wittgenstein distinguishes between what he calls 'I' or 'my' used as object and 'I' or 'my' used as subject (hereafter I_o and I_s respectively). He maps the distinction in the following way:

Examples of [I_o] are these: "My arm is broken," "I have grown six inches," "I have a bump on my forehead," "The wind blows my hair about." Examples of [I_s] are: "I see so-and-so," "I hear so-and-so," "I try to lift my arm," "I think it will rain," "I have toothache." One can point to the difference between these two categories by saying: The cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of an error, or as I should rather put it: The possibility of an error has been provided for. The possibility of failing to score has been provided for in a pin game. On the other hand, it is not one of the hazards of the game that the balls should fail to come up if I have put a penny in the slot. It is possible that, say in an accident, I should feel a pain in my arm, see a broken arm at my side, and think it is mine, when really it is my neighbor's. And I could, looking into a mirror, mistake a bump on his forehead for one on mine. On the other hand, there is no question of recognizing a person when I say I have toothache. To ask "are you sure that it's you who have pains?" would be nonsensical. Now, when in this case no error is possible, it is because the move which we might be inclined to think of as an error, a "bad move", is no move of the game at all. (1969, pp. 66-67, emphasis his throughout)

³ A caveat: the Blue Book consists of Wittgenstein's notes written in English and dictated to his students in 1933-34. And though he did send a copy of it to Russell, Rush Rhees reminds us that Wittgenstein did not consider it as work he intended to publish (see 1958, p. v).

Instances of I_s differ fundamentally from cases of I_o , then, in that with the former no room exists for errors of self-misidentification. Misidentification is precluded because I_s involves no recognition of any subject, and hence questions that one would normally ask to fix the identification of the subject in I_o or third-person cases prove nonsensical.⁴

To put the point another way, recall that we characterized self-misidentification in the previous chapter as one's willingness assent to "a is F," where a is a self-designating expression, while being disposed to deny that oneself is F. In these terms, Wittgenstein denies that it even makes sense to consider whether a subject could deny that she is F while willing to assent to a thought that she would express with the form " I_s am F." The "hazard" does remain, however, that a subject can mistake another for oneself in making self-ascriptions with I_o --Wittgenstein does consider this a possible "move of the game" of self-ascription--and thus one can indeed wonder in those instances who exactly is F.

⁴ At the risk of sounding pejorative, it does strike me as quite interesting that the list of examples Wittgenstein gives of 'I' used as subject line up rather neatly with Descartes' description of himself as a thinking thing in second Meditation: "A thing that doubts, understands, affirms, denies, is willing, is unwilling, and also imagines and has sensory perceptions" (1984, p. 19).

Wittgenstein's I_s/I_o distinction divides prima facie according to ascriptions that depend upon bodily observation and recognition, and those that do not. Accordingly, one is understandably tempted to conclude that the different uses of 'I' have distinct referents. On the strength of Wittgenstein's examples given above, I_s seems restricted to the realm of the mental, indicating that it must refer to something like a Cartesian ego or self. Ascriptions involving I_o are quite often bodily ascriptions, and the first-person pronoun used in this way appears to indicate or to single out the body the predicate is applied to, a body that may or may not be one's own. But this seems to imply that I_s refers to something conceivably independent of one's body, to something that can stake claim to it or that determines which body is one's own.⁵

However tempting this inference may be, Wittgenstein cautions against drawing it. Instead, he claims that the unique feature of I_s indicates that it does not function as a referring expression at all:

And now this way of stating our idea suggests itself: that it is as impossible that in making the statement "I have

⁵ Wittgenstein himself recognizes the pull of this construal. In "Notes for Lectures on 'Private Experience' and 'Sense Data'" (1993) he writes that "[i]t seems that I can trace my identity, quite independent of the identity of my body. And the idea is suggested that I trace the identity of something dwelling in a body, the identity of my mind" (p. 270).

toothache" I should have mistaken another person for myself, as it is to moan with pain by mistake, having mistaken someone else for me. To say, "I have pain" is no more a statement about a particular person than moaning is. ... The mouth which says 'I' or the hand which is raised to indicate that it is I who wish to speak, or I who have toothache, does not thereby point to anything. If, on the other hand, I wish to indicate the place of my pain, I point. And here again remember the difference between pointing to the painful spot without being led by the eye and on the other hand pointing to a scar on my body after looking for it. ("That's where I was vaccinated.") - The man who cries out with pain, or says that he has pain, doesn't choose the mouth which says it. (1969, pp. 67-68; emphasis his throughout)⁶

For Wittgenstein, then, referential expressions--much like pointing or gesturing, which are "led by the eye"--indicate or pick out their referents ostensibly by way of some observational link between subject and referent.

Ascriptions involving I_o do depend upon observation and are therefore subject to routine observational error and to the correlative mistakes of empty or misguided reference.

Those made with I_s , however, cannot fall into error, and

⁶ Again quoting from the Lectures (1993): "There is no difference, for me, between I and this ↗; and for me the word 'I' is not a signal calling attention to a place or a person" (p. 269). The arrow stands for a demonstrative gesture.

hence no reference or "picking out" must occur in such cases.⁷

Wittgenstein's resistance to Cartesian pressures may be admirable, but his alternative picture has several unwelcome consequences. Perhaps most strikingly, Wittgenstein's own examples seem to weigh against him, for raising one's hand to indicate a desire to speak does seem like a paradigm case of one's "pointing to" or indicating oneself. And in telling my doctor about my pain I certainly want her to be quite clear about which one of us

⁷ On several occasions, Wittgenstein attempts to clarify his claim regarding I_s (and what he often refers to as "how much truth there is in solipsism" [*Tractatus* 5.62]) by drawing an analogy between the physical eye and the distal visual field. The physical eye is not a constituent of the distal visual field but instead determines its boundaries. To describe what one sees one describes what's seen, and one's physical eye, since unseen, does not enter into this description. Similarly, Wittgenstein claims, the subject is not part of the world it limits: "A description of a sensation does not contain a description of a sense-organ, nor, therefore, of a person," and "Just as no [physical] eye is involved in seeing, so no Ego is involved in thinking or in having toothache" (1993, p. 100). When one experiences pain, one simply describes the pain as "it hurts"--no subject enters into such a description.

But this analogy confuses more than clarifies. Wittgenstein says little that is decipherable about what he means by "subject", and it's not at all clear how the analogy lines up generally. Presumably the subject corresponds to the physical eye, but what corresponds to the objects that make up the distal visual field? At times (*Tractatus* 5.631, e.g.) the compliment looks to be the physical world, but at others (e.g., 1993, p. 100) he seems to be concerned with what he calls "primary experiences." Most importantly, though, the analogy simply expresses Wittgenstein's conviction that no self-indication occurs with I_s. We have no reason to find the analogy applicable, and hence it provides no additional argumentative support. Gunderson (1970) also persuasively argues that even granting the applicability of the analogy, it does not license the radical conclusion Wittgenstein draws. For an extended discussion of Wittgenstein's treatment of solipsism, see Pears (1987).

requires medical attention, especially if I must distinguish myself from others to receive it.

But Wittgenstein wants to deny exactly this, that the first-person pronoun as it appears in ascriptions such as "I am in pain" serves to single out a particular person. His reasoning for this conclusion parallels his treatment of what he calls "Moore's paradox" (1975, Part II, p. 190). Moore's-paradox sentences are those that have the general form "p but I don't believe that p."⁸ Such sentences prove paradoxical because, as Moore (1942) himself noted, they seem absurd but not contradictory. One cannot say anything by asserting them, but--at least prima facie--they can be true.⁹

Wittgenstein finds these sentences particularly puzzling because they seem to evince a peculiar feature of first-person utterances. In Part II of the Investigations, Wittgenstein writes:

Think of the expression "I say", for example in "I say it will rain today", which simply comes to the same thing as the assertion "It will". "He says it will

⁸ These sentences are not limited to belief statements; analogues can be readily constructed for each of the attitudes--e.g., "Wish you were here, but I don't miss you," "Goats will be here shortly, but I don't expect any," and so on.

⁹ Controversy exists as to whether such statements are in fact outright contradictions. See, e.g., Adler (2002) and Shoemaker (1996).

...." means approximately "He believes it will". (1970, p. 192)¹⁰

Since they do have roughly the same performance conditions, one can and often does use either "It's going to rain" or "I think that it's going to rain" to convey one's meteorological convictions, for "the statement 'I believe it's going to rain' has a meaning like, that is to say a use like, 'It's going to rain'..." (1970, p. 190). If the meaning of a speech act were exhausted by the conditions for its use, then one might well conclude, as Wittgenstein does, that utterances of "p" and "I believe that p" have the same meaning and that assertions of mental attitude such as 'I believe', 'I expect', 'I wish', etc. are "merely ornamental pleonasms" (Rosenthal, 1986a, p. 156).

Similarly, the conditions of use for verbal and non-verbal expressions of pain overlap, and Wittgenstein regards the first-person pronoun as ornamental in these cases as well. In §404 of the Investigations Wittgenstein echoes his remarks from the Blue Book quoted earlier:

"When I say 'I am in pain', I do not point to a person who is in pain, since in a certain sense I have no idea who is." And this can be given a justification. For the main point is: I did not say that such-and-such a person was in pain, but "I am" Now in saying this I don't name any person. Just as I don't name anyone when I groan

¹⁰ Cf. "Moore's Notes," pp. 99-101.

with pain. Though someone else sees who is in pain from the groaning....

Groans aren't speech acts; one doesn't use them to say anything about anyone. If one's utterances of "I am in pain" were indeed used like groans, they would not qualify as self-ascriptions at all and would not specify a subject of predication as Wittgenstein claims. Another may respond to one's groans or verbal expressions of pain and run to one's aid, but she does so not because one tells her of one's pain.

Paul Horwich (personal communication) recommends against reading Wittgenstein as literally equating I_s utterances with groans. As evidence, he points to Wittgenstein's apparent response to that very accusation in §244 of the Investigations: "'So you are saying that the word "pain" really means crying?'--On the contrary: the verbal expression of pain replaces crying and does not describe it." But this remark simply provides additional evidence in favor of the point. If verbal expressions of pain do replace crying, presumably they do so in terms of their use--that is, budding English speakers learn to utter "I am in pain" on roughly the same occasions when they would have groaned in pain. And as we saw above, Wittgenstein holds that groans don't specify a particular

subject, which means that their verbal replacements must not do so either. Such expressions presumably don't mean crying because, as surrogate groans, they don't "mean" at all.¹¹

However, in assimilating groans and verbal expressions of pain involving I_s , as well as utterances of "p" and "I believe that p", Wittgenstein runs together a crucial distinction. As David Rosenthal (1986a, 1993, 1995, 2002) has noted, Wittgenstein conflates expressing and reporting mental states. A particular speech act expresses a mental state just in case it has the same propositional content as that state and its illocutionary force corresponds to the state's mental attitude. When I assert, for example:

(11) Goats like grass

I express my belief that goats like grass, and

(12) Goats will be here shortly

expresses my expectation regarding the imminent arrival of goats.

¹¹ Even reading the passage as charitably as possible, in accepting the impossibility of misidentifying oneself when 'I' is used as subject, Wittgenstein seems to set up a false dilemma. He only allows for explanations of this feature of I_s in terms of having one's Cartesian ego as its referent or in terms of its not being a referential expression at all--no middle ground exists where instances of I_s could refer to something other than a Cartesian ego.

I can also convey my thoughts by reporting them. In asserting

(13) I believe that goats like grass

I report my belief about goats, and in asserting

(14) I anticipate that goats will be here shortly

I report my expectation.¹² In neither case do the reports have the same propositional content as the state they report--I use them to say something not about goats per se but about what thoughts I have about them. And whereas (12) has the illocutionary force of expectation, the report of that expectation, like all reports, has the force of an assertion.

On this picture, the air of mystery surrounding Moore's-paradox sentences results from a unique combination of performance conditions and truth conditions. Such sentences have no coherent performance conditions, for in asserting them one would both express a particular thought and deny that one had it. The two conjuncts do differ in truth conditions, however, and the content of the sentence as a whole is therefore not contradictory. Rosenthal writes:

¹² Reports are themselves expressions of mental states, and reports of those states add another level of propositional attitude--e.g., "I believe that I believe that goats like grass." All reports are expressions, but not vice versa.

To explain the divergence manifest in Moore's paradox between truth conditions and conditions for the coherent performance of speech acts, we must distinguish expressing mental states from reporting that one is in them. ... If expression were not distinct from reporting, the first conjunct would both express and report one's belief, whereas the second conjunct would still deny that any such belief exists. So on this construal the sentence would be contradictory which it plainly is not. Accordingly, a correct explanation of Moore's paradox is impossible unless we recognize that reporting a mental state is distinct from expressing it. (1986a, p. 155)

The truth of (11) depends upon the gastronomic preferences of goats, whereas the truth of (13) depends upon the state of my beliefs. It might be the case that I don't believe that goats like grass when in fact they do, but any attempt I might make to assert

(15) Goats like grass but I don't believe that they do fails as a speech act.

If, as Wittgenstein claims, assertions such as "I am in pain" were of a piece with groans or interjections such as "Ouch!", Moore's paradox would be contradictory as well as absurd. For if mental state expressions and their reports were indeed indistinguishable, their truth conditions would be identical as well, leaving Moore's-paradox sentences tantamount to sentences of the form "p

and not p." As we saw above, however, the truth conditions of the two conjuncts do clearly differ, and only by respecting this difference can we capture the peculiar features of Moore's paradox.

The difficulties run deeper. Exclamations of pain such as "Ouch!" or groans do not have truth conditions at all;¹³ they have no propositional content and don't purport to predicate properties of anything. Yet when I assert "I am in pain," I seem to say something truth evaluable, true just in case I am in pain and false otherwise. If assertions like "I am in pain" were equivalent to "Ouch!", in expressing the former I have said nothing that is either true or false. This in turn means that certain states and their corresponding speech acts cannot fund the expected logical entailments. Speech acts such as "He is in pain" license the inference that someone is in pain, but for Wittgenstein those involving I_s such as "I am in pain" do not. Likewise, one could not literally deny the truth of another's first-person assertion that she is in pain. This should strike us as strange, especially given that the puzzle of distinguishing nonaccidental from accidental

¹³ Wittgenstein could argue that since moans have no truth conditions, Moore's paradox sentences are not strictly contradictory. Thorny issues regarding expressing and reporting sensory states do lie in wait here, but suffice it to say that Moore's paradox-sentences are not contradictory because they can be true. See Rosenthal (1993).

self-ascriptions proved initially troubling because thoughts and their corresponding speech acts such as "I am in pain" and "R.M. is in pain" share truth conditions but have divergent entailment relations.

In addition, more complex but fairly commonsense reports of beliefs that appear to draw upon both uses of 'I' become utterly mysterious. Consider

(16) I believe that I have grown six inches.

How are we to understand (16)? The second token of 'I' used to report my belief regarding my growth spurt is I_o . The utterance as a whole, in contrast, represents an instance of I_s . But presumably groans cannot report self-ascriptions that have truth-conditions, which means that logical relations are lost here as well. We normally take (16) to entail assertions such as

(17) I believe that someone has grown six inches

and

(18) He believes that he has grown six inches

said of me by another. Though the surface grammar of an utterance may differ markedly from its deeper structure, denying the validity of inferences from first-person ascriptions to their existentially generalized counterparts forces us to give up too much.

In light of this rich evidence, why doesn't Wittgenstein sign on to the expressing/reporting distinction and thereby avoid the fairly drastic consequences that follow from ignoring it? Rosenthal (1986a) diagnoses Wittgenstein's error as resulting from his appreciation of the use of a particular speech act at the expense of its truth conditions.¹⁴ For "if we take the use of linguistic expressions as central to their semantic character, it will be difficult not to assimilate reporting mental states to expressing them" (Rosenthal 1986a, p. 156).

No doubt this consideration largely explains his conflation. Still, confusing use and truth-conditions does not fully explain Wittgenstein's denial that I_s specifies the ascribing subject. For if expressions and reports do have roughly the same use, and therefore roughly the same meaning, why not think that "It's raining" or "Ouch!" actually reports one's mental state--that in uttering either one specifies a subject who has a belief about the weather or who has pain, as one does with the speech act "I believe that it's raining" or "I am in pain" traditionally construed? Wittgenstein seems alive to this question:

¹⁴ Cf. Investigations §43: " For a large class of cases--though not for all--in which we employ the word meaning it can be defined thus: the meaning of a word is its use in the language game."

Consider the misbegotten sentence "It may be raining, but it isn't".

And here one should be on one's guard against saying that "It may be raining" really means "I think it'll be raining." For why not the other way round, why should not the latter mean the former? (1970, II, p. 192)

If "It's raining" and "I believe that it's raining" do roughly share performance conditions, what decides which should be regimented as the other?

We're never explicitly told why we should be on our guard, but another concern probably contributes to the conclusion that reports are actually expressions in grammatical disguise: Moore's paradox only arises for first-person, present-tense sentences.¹⁵ Sentences such as "Goats like grass but I didn't believe it" or "Goats like grass but he doesn't believe it" are not contradictory and can be coherently asserted. Asserting "p" expresses the mental attitude one takes towards a content, but in reporting that thought one not only asserts which mental attitude one takes, one also specifies who takes it. Absurdity results in Moore's-paradox cases in large part because in asserting them one would both express one's

¹⁵ Wittgenstein himself points out that tense plays a role in the puzzle: "the statement 'I believe it's going to rain' has a meaning like, that is to say a use like, 'It's going to rain', but the meaning of 'I believed then that it was going to rain', is not like that of 'It did rain then'" (1970, p. 190).

attitude towards a content and deny that oneself was the thinker or speaker who holds that very attitude towards that very content.¹⁶

Taken together, Wittgenstein's convictions regarding meaning and the possibility of first-person self-misidentification lead to his troubling account of nonaccidental self-ascription. For, according to Wittgenstein, errors of identification can always occur in specifying or picking out an object. Consequently, if either expressions or reports were used to specify who exactly believes, hopes, or expects such-and-such, that specification could go awry as it can when one, say, mistakes another's reflection for one's own in a crowded mirror. But Wittgenstein maintains that ascriptions involving *I*_s preclude errors of self-misidentification, which means that tokens of *I*_s--including those figuring in mental state reports--must not be used to specify anyone. Similar use may encourage conflating expressions and reports, but the impossibility of self-misidentification dictates that utterances involving *I*_s should be regimented as expressions in which the expressing subject does not

¹⁶ Descartes' cogito provides an additional example. The sentence "I don't exist" was true and will be true again when indexed to me, but it has no conditions of assertability. Asserting it defeats it as a speech act.

enter into their content, regardless of their surface grammar. The subject thus becomes a casualty of the expressing/reporting conflation--it is made to disappear altogether so that it cannot be missed.¹⁷

Whatever one may think about the importance of use in determining the proper regimentation of linguistic expressions and the illustrative powers of Moore's paradox, claims about the impossibility of self-misidentification need further support. Without surface grammar to guide us, how are we to tell whether 'I' is used as subject or as object on a particular occasion? The most natural way to draw the distinction between nonaccidental and accidental self-ascriptions is, as Chisholm suggests, according to those that one would express with the first-person pronoun or indirect reflexive and those that one would express with other self-designating expressions or the ordinary reflexive. Wittgenstein effectively denies that self-ascriptions wear their natures on their grammatical sleeves, however, and maintaining a stark contrast between accidental and nonaccidental self-ascriptions requires some relatively clear means of making out the two categories.

¹⁷ Cf. his similar argument (1970, §246, pp. 221-222) that one cannot truly be said to possess self-knowledge since one can't know anything without the possibility of not knowing and it doesn't make sense for one to doubt, e.g., whether one is in pain.

Drawing the distinction as Wittgenstein does by limiting self-ascriptions that preclude errors of self-misidentification to those ascribing mental states and properties is controversial. Gareth Evans (1982), for example, finds that certain bodily ascriptions also leave no room for questions regarding to whom a property is ascribed. One's kinesthetic and proprioceptive system provides a stream of information regarding the state of one's body, the position of limbs, skin and joint tension, bodily feedback during motion, etc.¹⁸ Evans argues that kinaesthetic and proprioceptive information gained about one's body via these systems in the normal way "appears to give rise to judgments which are immune to error though misidentification" (p. 220). Just as it makes no sense to ask "Some is in pain, but is it I?", questions such as "Someone's legs are crossed, but is it my legs that are crossed?" or "Someone is hot and sticky, but is it I who am hot and sticky?" are equally out of place. Similarly, José Luis Bermúdez (1998) argues that proprioception provides information about the state of one's body and only one's body. Questions regarding whose body is proprioceptively perceived cannot arise, for "[i]t follows from the simple

¹⁸ For a more extensive list of the informational systems that constitute somatic proprioception, see the general introduction to Bermúdez, Marcel, and Eilan (1995), pp. 13-15.

fact that I somatically proprioceive particular bodily properties and introspect particular psychological properties that those bodily and psychological properties are my own" (Bermúdez 1998, p. 147).

These arguments have their own opponents, of course, and we will revisit their finer points in the following chapter. But for the moment we should note that this difference of opinion emphasizes an important shortcoming in Wittgenstein's discussion. He provides no substantive reason for accepting the impossibility of self-misidentification in the cases he favors. Perhaps certain self-ascriptions of one sort or another do not "leave room" for one to misidentify oneself, but he gives us no principled way to regiment self-ascription tokens as either accidental or nonaccidental apart from any intuitions we may have about which ones are subject to the relevant sort of error. But our present task is to elucidate and to make sense of those very intuitions and to get clearer about what exactly is the relevant sort of error. What we need is insight into what Wittgenstein and others mean by the claim that one simply cannot mistake another for oneself in certain cases and how that claim figures into one's ability to self-ascribe properties.

§3. Nonaccidental Self-Ascriptions and Guaranteed Self-Reference

G. E. M. Anscombe's (1975) influential discussion of the first person holds promise for elucidating the elusive assumption regarding self-misidentification so central to Wittgenstein's picture of self-ascription. She also subscribes to the view that 'I' isn't used to specify or to refer to a subject, arguing that, like other familiar statements involving purely grammatical placeholders, tokens of the first-person pronoun in nonaccidental self-ascriptions are expletive words, or words that "fill a syntactic gap but make no contribution to the sense of the sentence" (Katz 1988, p. 119). In asserting "It's raining," for example, I say something true or false according to current weather conditions, but the truth of my assertion doesn't seem to depend upon the reference of 'it'. Similarly, she holds that one's utterances of "I am in pain" actually say something similar to "It hurts," which is true or false according to my current pain state. 'It' here does not even purport to refer. As a result, first-person self-ascriptions amount to "unmediated conceptions (knowledge or belief, true or false) of states motions, etc., of this object here..." (Anscombe 1975, p. 62).

Anscombe's overall strategy for reaching this conclusion resembles Wittgenstein's in many respects. She too argues that one can speak of or refer to oneself while failing to realize that one is doing so. Uses of 'I', however, could have nothing less than guaranteed reference if they did in fact refer. Since no referring expression could secure this strong a referential guarantee, the first-person pronoun either requires a special kind of permanent and persisting referent such as a Cartesian ego or it must not serve to refer at all. Choosing the former option leads to a host of historically intractable problems. Hence, Anscombe concludes that we should take the latter, Wittgensteinian one in which "'I' is neither a name nor another kind of expression whose logical role is to make a reference, at all" (1975, p. 60, emphasis hers).

The premise that interests us here is the middle one—namely, that the first-person pronoun could have nothing less than guaranteed reference if it were to refer at all. Anscombe claims that this guarantee must be two-fold: the referent of 'I' must exist and it must be the thing that the speaker using it intends to refer to. That is, the speaker cannot "take the wrong object to be the object he means by 'I'" (1975, p. 57).

Anscombe elucidates what she sees as the difference between these types of guarantee by imagining a name that functions referentially similarly to 'I'--should 'I' refer--in that it is an expression each speaker uses to refer only to herself. She writes:

Imagine a society in which everyone is labeled with two names. One appears on their backs and at the top of their chests, and these names, which their bearers cannot see, are various: 'B' to 'Z' let us say. The other, 'A', is stamped on the inside of their wrists, and is the same for everyone. In making reports on people's actions everyone uses the names on their chests or backs if he can see these names or is used to seeing them. Everyone also learns to respond to utterance of the name on his own chest and back in the sort of way and circumstances in which we tend to respond to utterance of our names. Reports on one's own actions, which one gives straight off from observation, are made using the name on the wrist. Such reports are made, not on the basis of observation alone, but also on that of inference and testimony or on other information. B, for example, derives conclusions expressed by sentences with 'A' as subject, from other people's statements using 'B' as subject. (1975, p. 49)

Anscombe's 'A'-users uses of 'A' are guaranteed to refer in the first sense in that no self-ascription involving 'A' could occur without an 'A'-user to token it. However, an 'A'-user might mistake another's wrist for her own and ascribe a property with 'A' as subject on that basis. Uses of 'A' therefore fail to be guaranteed in the second sense.

Some peculiarities in this example are worth noting. As she describes it, 'A'-users ascribe properties to themselves using sentences including 'A' as subject grounded in observation, inference, or "other information". Self-ascriptions based on one's own actions are given "straight off from observation ... using the name on the wrist," which seems to imply that 'A'-users must witness some bit of bodily behavior, locate the name on the wrist that falls under their observation in a "characteristically limited and also characteristically privileged" way (p. 49), and then ascribe the relevant property.¹⁹ She does allow that an 'A'-user can eventually learn that others refer to her by the unique name tattooed on her back and chest "in the sort of way and circumstances in which we tend to respond to utterance of our names." Anscombe's speakers can also infer from an interlocutor's use of these names to a self-ascription made with 'A'.

What Anscombe does not seem to leave room for is any "direct" self-ascription with 'A' free from observation or conscious inference. But presumably whatever capacities her speakers possess that allow them to respond to utterances of their unique names as we do to ours should equip them with the ability to self-ascribe with 'A'

¹⁹ O'Brien (1994) is similarly puzzled by Anscombe's example.

without observing wrists of any sort or without consciously inferring that they are identical to the individual being referred to. Perhaps when someone calls out "Roblin is wanted in the kitchen" from the next room I must appeal to--or at least must be disposed to appeal to--my belief that I am Roblin in order to infer that I am wanted in the kitchen. Nevertheless, such an inference certainly need not be conscious. Similarly, one can self-ascribe properties with utterances of one's name without consciously working through the inference that one is the referent of the name. It's "automatic", we're inclined to say about our own case. If 'A'-users are indeed supposed to share our ability to respond to our given names, not all of their self-ascriptions need involve conscious inference. We must be on our guard against questions being begged.

Leaving the peculiarities of her example aside for the moment, Anscombe claims that one notable difference between 'A' and 'I' is that unlike uses of 'I', uses of 'A' do not express self-consciousness on the part of their users. We should resist the temptation to unpack self-consciousness as consciousness of a self, however, either as a particular thing or as a unique presentation of oneself to oneself. Anscombe believes that thinking of self-consciousness in this way is "strictly nonsensical,"

for it raises the insoluble problem of how to guarantee that one has got hold of the right self when using the first-person pronoun. Instead, she argues that the only sensible understanding of self-consciousness is "consciousness that such-and-such holds of oneself" where 'oneself' is the indirect reflexive (1975, p. 51). As the indirect reflexive, 'oneself'--like 'himself'--gives us no added traction on the question of self-reference since it is nothing other than the indirect equivalent of 'I'. We only end up back where we began.

The main dissimilarity between the two expressions, one adumbrated above in Anscombe's characterization of self-consciousness, is that uses of 'I' are guaranteed against referential errors. She intends her 'A'-user example to demonstrate that no proper name can provide the relevant senses of referential guarantee necessary to capture the features unique to uses of 'I'. She writes:

It seems clear that, if 'I' is a "referring expression" at all, it has both kinds of guaranteed reference. The object an 'I'-user means by it must exist so long as he is using 'I', nor can he take the wrong object to be the object he means by 'I'. (The bishop may take the lady's knee for his, but could he take the lady herself to be himself?)" (1975, p. 57)

Though the grammatical role of the first-person pronoun may make it look like a referential expression, if it were a

sort of name like 'A', then a speaker using 'I' could in fact be mistaken about its referent. But as much as the bishop might like to take the lady for himself, one's uses of 'I' do not seem to allow for that possibility.

Not only do names fall short of supplying the additional level of guarantee, Anscombe believes that no referring expression can meet this demand. She argues that definite descriptions and demonstratives, like names, refer in virtue of a speaker's possessing a conception associated with the term that "latches onto" a particular object. No conception seems able to guard against one's taking another object to be oneself; one always seems able to refer to something other than what one takes oneself to be referring to.

If guaranteed reference cannot be explained in terms of how 'I' refers via some conception, perhaps it can be accounted for in terms of what it refers to. It seems that only a Cartesian ego could ground one's use of 'I', for it alone is the kind of thing that is, as Anscombe puts it, "freshly defined with each use of 'I'" and that "always remained in view so long as something was being taken to be I" (1975, p. 57). One's body, for example, need not be constantly "present" or "in view" to one. Intuitively it seems that one could find oneself in a state of sensory

deprivation--cut off from all exteroceptive and proprioceptive sensory information about one's body--yet still be able to think first-person thoughts. Given that one's body cannot sufficiently guarantee first-person reference, 'I' refers only if "Descartes was right about what the referent was" (Anscombe, 1975, p. 58).

Like Wittgenstein, Anscombe inveighs against adopting Cartesian metaphysics and the intractable problems it gives rise to. Even apart from its traditional problems such as the interaction between mind and body, if one's use of 'I' stood for one's immaterial soul, we would still face the problem of whether all of one's uses of 'I' referred to the same thing. We therefore find ourselves driven towards the Wittgensteinian conclusion that 'I' doesn't refer:

Getting hold of the wrong object is excluded, and that makes us think that getting hold of the right object is guaranteed. But the reason is that there is no getting hold of an object at all. With names, or denoting expressions (in Russell's sense), there are two things to grasp: the kind of use and what to apply them to from time to time. With 'I' there is only the use. (p. 59)

We cannot avail ourselves of any type of referring expression to explain nonaccidental self-ascriptions, and we certainly do not want to rely upon Cartesian immaterial substances. Having exhausted our best attempts at

construing the first-person pronoun as a referring expression, we arrive via reductio at the conclusion that "'I' is neither a name nor another kind of expression whose logical role is make a reference, at all" (1975, p. 60).

Ironically, Anscombe assures us that the many difficult philosophical problems traditionally associated with the first person can be avoided altogether if we abandon the Cartesian assumption that 'I' must be a referring expression. Her conclusion that the first-person pronoun isn't an expression of singular designation, however, engenders its own disputes that seem equally likely to be "self-perpetuating, endless, irresoluble" (Anscombe 1975, p. 60). Cartesian metaphysics may be troubling, but its offenses against our philosophical sensibilities may well be matched by Anscombe's alternative.

One major offense is against English grammar. The first-person pronoun may not be a type of name, but Anscombe errs in her argument that considering it a pronoun tells us little about its referential status. To use the example Anscombe herself gives, 'he' in the sentence "If anyone says that, he is a fool" behaves like a variable and not a term of singular designation. However, as Jerrold Katz observes, 'I' does not function analogously in any of

its uses. Substituting 'I' for 'he' in the sentence above returns "If anyone says that, I am a fool," which does in fact denote a particular speaker. According to Katz, Anscombe's own example demonstrates that "'I' does not function via an anaphoric link back to a quantified subject, but initiates reference on its own" (1988, p. 123).

Grammatical evidence can also be marshaled against the general position that 'I' is an expletive word.²⁰ Certain syntactic constructions do permit an expletive reading of the subject placeholder such as

(16) It is wet.

Others, however, require that 'it' be a referential term, such as

(17) It is pleased to be wet

for 'it' must be the subject of both 'is pleased' and 'is wet'. By analogy, the first-person pronoun in sentences such as

²⁰ Georg Christoph Lichtenberg is perhaps best known for this view as an objection to Descartes' cogito, that "I think" should be understood as "It thinks." Geach (1957) and Strawson (1959, pp. 99-100) agree, though they both think that we do use 'I' in a referential capacity when we communicate our self-ascriptions to others. Nietzsche (1966), as one might expect, goes one step further, claiming that even in limiting the cogito to the "subjectless" "It thinks" already involves an interpretation forced by the demands of grammar. Wittgenstein thought favorably of Lichtenberg's view as well (see 1993, pp. 100-101).

(18) I am pleased to be wet

must also be referential for the same reason. Similarly, conjunctions involving pro-forms such as 'too' and 'so' observe syntactic equivalence in the subject position when the conjuncts share a verb phrase (Katz, 1988, p. 126). The 'it' in a sentence such as

(19) Steve is wet and it is too

must be read referentially since 'it' borrows 'wet' from the first conjunct and so must be syntactically equivalent to the referential term 'Steve'. If 'I' always functioned as an expletive word, sentences of the form

(20) Steve is wet and I am too

would violate subject equivalence. (20) doesn't violate equivalence, indicating that 'I' is as referential as 'Steve'.²¹

Performatives provide additional pressure against an expletive interpretation of 'I'. Performatives are illocutionary acts executed with utterances employing certain verbs of illocutionary force, and they are generally framed in first-person singular present tense (Austin 1962, Vendler 1972). One can promise to play horseshoes on Friday, for example, with the statement "I

²¹ This paragraph, like the one before it, draws heavily upon Katz (1988), Chapter VIII.

promise to play horseshoes on Friday." Its second- and third-person variants, however, do not perform particular acts but rather report them.²² In arbitrating a disputed weekend schedule I might assert that "He promises to play horseshoes on Friday." But in doing so I have not thereby committed myself to any activities. And the subject term in the sentence "It promises to play horseshoes on Friday" can only have a referential reading. This isn't surprising given that performatives are illocutionary acts, and acts require agents to perform them.

We raised a similar tense and person objection in connection with Moore's paradox in the last section, and it applies with equal force here as well. If with 'I' "there is only the use" as Anscombe says, and it is not used to

²² Things are a bit more complicated, of course. First-person performatives phrased in tenses other than the present usually have the illocutionary force of a reports. I do not make a promise by stating "I have promised to play horseshoes." An exception is future tense-- 'will' combined with the first-person often serves as a route to commitment as in "I will dump the trash." Future progressive statements such as "I will be arriving on the D train" likewise involve commitment, but they don't fit neatly into the performative category because they don't necessarily single out a moment at which an illocutionary act occurs. Arguably, an exception exists to the first-person rule of performatives as well. 'Shall' as used in legal documents imposes a duty upon the named party, regardless of person, as in "Tenant shall pay landlord upon occupation of premises." But in these cases the third person serves the specifically referential purpose of specifying exactly which party is committed. Performatives can also be made in the first-person plural, as when a team confidently offers "We swear we will do better next game." Both of these latter examples require further elucidation, but the main point in the text stands that at least certain performatives such as promising require specifying who has promised. See Vendler (1972) for a detailed discussion of these issues.

ascribe properties to a particular subject, then Moore's-paradox sentences become outright contradictions. Anscombe has little to say about how we should handle what she calls "Cartesianly preferred" thoughts such as thinking or experiencing (1975, p. 63), but presumably 'I' does not refer to a subject in those cases as well. This means that Moore's-paradox sentences of the form "p but I don't believe that p" amount to "p but not p," a result which effectively closes off any plausible avenue of explanation of the paradoxical but not contradictory character of such sentences. As we saw above, a satisfactory account requires respecting their first-person, present-tense nature.

Questions regarding the logical relations supported by first-person ascriptions help to evince a few more of the perplexities lurking in Anscombe's positive account. Anscombe wisely doesn't want to deny that these ascriptions are truth evaluable. In her view the thought that one would express with "I am standing" is true or false according to whether "this object here" or "this thing here" is standing. Presumably these equivalents entail that something is standing and can be contradicted by the claim that nothing is standing. Does "I am standing" uttered by Anscombe entail "She is standing" uttered of her

by another? Again, presumably--one can correctly infer that "He is dangerous" from another's warning that "This suspect here is dangerous." No special problems of entailment arise with demonstrative statements in general.

This straightforward reading won't do, however. Cashing out first-person ascriptions in terms of equivalents phrased with demonstratives can at best approximate what Anscombe wants to capture about the first person. After all, she claims that one's utterance of "I am this thing here" is not an identity (1975, p. 61). More importantly, she herself argues that 'I' cannot be assimilated to demonstratives--they are referential expressions that require a conception by which they reach out to their referents, and for that reason they always remain subject to referential errors. Accordingly, one may be wrong about which object the demonstrative phrase "this object here" picks out. Demonstrative equivalents might be best understood as describing the conditions under which a first-person ascription is verified, whereas self-ascriptions themselves are "unmediated conceptions (knowledge or belief, true or false) of states, motions, etc., of this object here..." (1975, p. 62).

Whatever grip we may have on what "unmediated conceptions" are exactly and how they can be true or false

quickly loosens in the face of Anscombe's treatment of sensory deprivation. She imagines a case in which one receives no sensory information whatsoever about one's body, and one is unable to speak. She finds it intuitive that one could still think first-person thoughts in this case, perhaps telling oneself "I won't let this happen again!" One could even doubt whether one had a body, entertaining the possibility that "there is then nothing that I am" (1975, p. 62).

How are we to understand "I won't let this happen again!" or how it could occur to one to think "There is nothing that I am"? One's ideas of actions and the like survive sensory deprivation, Anscombe believes, so apparently one could still think about moving and acting. But if "I won't let this happen again!" is to be understood as "this object here won't let this happen again," and sensory deprivation prevents 'this object' from latching onto anything as Anscombe admits, we reach a muddle. As Kenny (1984) nicely puts it, "if I no longer have a body, then I no longer exist, as Professor Anscombe explicitly concedes. And if I do not exist, then I cannot be making resolves either" (p. 85).²³

²³ Kenny (1984) objects that Anscombe's reconstrual of first-person ascriptions as "this object here of whose action this idea of

We can avoid the vexing suite of problems caused by Anscombe's no-reference view by resisting her radical conclusion. Her argument in favor of it depends upon the premise that a competent user of the first-person pronoun cannot "take the wrong object to be the object he means by 'I'" (1975, p. 57), if it were to refer at all. We can disarm her argument by showing this premise to be unacceptable or explainable in other terms that do not conflict with the possibility of first-person reference.

We first need to be clear as to just what kind of error this condition proscribes. Anscombe considers it a referential guarantee, so presumably one cannot erroneously refer to another in attempting to self-ascribe properties with the first-person pronoun. As she describes it, referential errors result from mistakes of identification, which simply cannot occur with competent uses of the first-person pronoun. 'A'-users can apparently fall into this sort of error when self-ascribing properties with 'A':

Of course, a man B may sometimes make a mistake through seeing the name 'A' on the wrist of a man whose other name is after all not inaccessible to B in the special way in which his own name ('B') is. (1975, p. 49)

action is an idea ..." relies upon private ostensive definition to individuate ideas. He considers this error, made explicit by her discussion of sensory deprivation, as clear evidence of "the power of Cartesian ideas to survive and flourish in the most hostile of climates" (p. 78).

As a result, "the 'A'-user would not be immune to mistaken identification of someone else as 'A'" (1975, p. 57).

Let's consider an 'A'-user's error in a bit more detail. C's wrist (let's say) falls under B's observation in the "characteristically special way" that B's wrist normally does. B then reports "A is F" on that basis, but she is mistaken--C is in fact F and B is not. How do we sort out the reference here? Does B's use of 'A' refer to C? Has B said something false of herself or true of C? Anscombe would certainly agree that B has mistakenly identified who or what was F in this case, for she has taken another to be herself. But does this misidentification lead B to refer to someone other than herself?

Though she never speaks to this question directly, Anscombe apparently would accept that B has said something true of C and that misreference does occur in cases of misidentification in general--the guarantee we want to explain is a referential guarantee after all. This doesn't seem to square with the purpose of her 'A'-user example, however. The example purports to answer the question "is it really true that 'I' is only not called a proper name because everyone uses it only to refer to himself?" (1975, p. 49, emphasis mine). The term 'A' is supposedly "a clear

case of just such a name." But if 'A'-users can misidentify the subject of an ascription, and if misidentification leads to misreference, then 'A' simply isn't a term only used by a speaker to refer to herself. It can be used to refer to someone else.

Perhaps that is Anscombe's point, that no referring term could play such a role. In setting up her example she avoids characterizing the performance conditions for 'A' in terms that suggest any sort of self-recognition on behalf of 'A'-users. Though she defines 'A' as the term used to "report on one's own actions," we are told that an 'A'-report is "one's own" only insofar as it issues, for example, "from the mouth of B on the actions of B" (1975, p. 49, emphasis mine). And whereas reports from B made with 'A' are "prima facie verified by ascertaining whether B did it..." (p. 49), instances of misidentification require that we go beneath the face of things to determine truth and falsity.

Uses of 'I' differ fundamentally in that they do not allow for a similar gap between the reporting mouth and the subject whose states and actions are reported. Anscombe writes:

If you are a speaker who says 'I', you do not find out what is saying 'I'. You do not for example look to see what apparatus the

noise comes out of and assume that that is the sayer; or frame the hypothesis of something connected with it that is the sayer. If that were in question, you could doubt whether anything was saying 'I'. As, indeed, you can doubt whether anything is saying it out loud. (1975, p. 56)

'A'-users must do exactly this--they must look to see who (or what) is behaving in such-and-such a way, and then ascribe properties on the basis of that observation. Accordingly, B can erroneously utter "A is F" upon observing C's wrist, effectively issuing a report from the mouth of B on the actions of C. Since B reports C's action instead of her own, she says something true about C and not something false about herself. Such are the consequences of referential terms corrupted by misidentification. Uses of 'I', in contrast, do not seem to rely upon any sort of observation. The speaker or thinker who uses 'I' is not in the same position as the hearer. When one hears an utterance involving the first-person pronoun, a question can arise as to who is the speaker or thinker. Such doubts about one's own use of 'I' does not seem possible.

But this description is just as true of one's common uses of one's own name. When I refer to myself as Roblin, I do not "look to see what apparatus the noise comes out of" and assume that that is Roblin. We can imagine cases in which I fail to recognize that I am Roblin, but we can

just as easily imagine cases in which I do. And in those cases, I don't consciously appeal to the premise or hypothesis that I am Roblin. Perhaps I do indeed make a nonconscious inference involving that premise, but even in that case it will seem to me that my being Roblin is unmediated and nonobservational, and that doubts about who is saying 'Roblin' cannot arise.

More importantly, there is a sense in which one can mistake the object meant by 'I' yet still succeed in referring to oneself. Consider the errors Wittgenstein appeals to to motivate the difference between I_0 and I_s , ones which bear a striking resemblance to those possible with uses of 'A' that Anscombe has in mind. To take one example, he imagines that I look into a mirror and mistake another's bumpy forehead for my own. If I then assert "I have a bumpy forehead," I say something false about myself even in mistaking another for myself. I do mean to refer to the person whose forehead is bumpy, but I believe that that person is myself.²⁴ 'I' functions as a token reflexive, and competent uses of it always refer back to their tokeners. If 'A' is to mirror the uses of 'I', an 'A'-user who ascribes a property with 'A' in virtue of

²⁴ Taschek (1985) offers a similar example that involves mistaking a photograph of another to be a picture of oneself.

mistaking another's wrist for her own nevertheless self-ascribes that property, albeit falsely. To be like 'I', 'A' must serve as a dedicated device of reflexive reference even in cases of misidentification.

Self-reference with the first-person pronoun even survives radical error about who one is. Imagine that I have come to believe that I am Freud and that a large range of my behavior is supported my delusion. If asked who I am I respond that I am Freud, and I refuse to answer to anyone who does not address me by that name. I incessantly attempt to analyze the dreams of those around me, and I am quick to assert that Civilization and Its Discontents was some of my best work. Needless to say, I am not Freud. Even in this extreme case of misidentification my first-person expressions of false Freud beliefs such as "I invented psychoanalysis" say something false about me and not something true about the actual Freud. Moreover, I do intend my use of the first-person pronoun to refer to Freud, and when hear others speak of Freud I believe that they are speaking of me. I am wrong both about who I am and who exactly is Freud.

Anscombe's referential guarantee, and with it her larger conclusion that 'I' isn't a referring expression, doesn't hold up under scrutiny. Still, she does seem to be

correct in appealing to the language of guarantee to describe uses of the first-person pronoun. Cases such as those discussed above indicate that we must make take a more nuanced approach to misidentification. Rosenthal (2003) agrees, arguing that we need to distinguish between identifying who it is that is F and whether it is I who is F. Though one may indeed be mistaken about who or what one is--which names and descriptions properly pick one out and which properties one instantiates--one nevertheless seems quite secure in one's belief about whether it is oneself who believes himself to be, for example, Freud.

We have a fairly good grasp of what it means for one to identify who it is that is F--one does so in virtue of possessing and deploying a battery of descriptive information about who is F. The latter half of the distinction, however, remains mysterious. Accordingly, Wittgenstein and Anscombe have brought us closer to the understanding we seek regarding the claim that misidentification is impossible in making certain self-ascriptions. We now face two interrelated questions: Specifically, what does it mean to say that I cannot be mistaken whether it is I who is F, and what does this special kind of freedom from error have to do with reference to oneself?

§4. Immunity to Error Through Misidentification Relative to the First-Person Pronoun

Our overarching goal in this chapter is to understand and to evaluate the claim that one cannot misidentify oneself when making nonaccidental self-ascriptions. Wittgenstein and Anscombe rely upon this assumption to argue that in few of its uses does the first-person pronoun specify oneself. If 'I' were to refer, the argument goes, then uses of the first-person pronoun could not be guaranteed against misidentification. Since one cannot mistake another for oneself when using 'I', such uses must not be referential. However, we saw that Wittgenstein offered little insight as to why we should accept this key claim and that Anscombe erred in thinking that misidentifying oneself, at least in one form, is incompatible with self-reference.

Nevertheless, we have come a step closer to meeting our larger goal in that we can now more precisely formulate the freedom from misidentification that certain self-ascriptions seem to enjoy: One can indeed be wrong about who one is, but even when mistaken in this way one still can't be wrong about whether or not it is oneself who has a particular property.

How are we to understand this apparent freedom from misidentification, and how does it relate to whether the

first-person pronoun is used to refer to oneself? Sydney Shoemaker attempts to explain this freedom in a way that does not preclude 'I' from being a referring expression. He believes that Wittgenstein has succeeded in singling out a puzzling class of self-ascriptions in need of a special account--namely, those made with I_s . To understand the mysteriousness of the first person that Wittgenstein and Anscombe are sensitive to that "underlies the perennial attractiveness of such unacceptable views about the self and self-reference" is not to agree with them, however (1968, p. 556).²⁵ For "nothing seems clearer than that in all first-person statements... the word 'I' functions as a singular term or singular referring expression" (1968, p. 555). Their mistake lies in thinking that self-reference is incompatible with the impossibility of self-misidentification. If they were in fact incompatible, Shoemaker claims that we would be at a loss to explain how a subject can ascribe properties generally and refer to anything at all.

²⁵ Shoemaker (1963) admits the pull of this temptation himself: "It can easily seem that in the actual making of first-person statements, or at any rate of first-person 'experience' or 'psychological' statements, the notion that the word 'I' refers to something plays no part at all; for me to be entitled to say 'I see that it is raining' it seems no more necessary that I should observe, or be able to identify, something designated by the word 'I' than that I should observe, or be able to identify, something designated by the word 'it'" (p. 10). Succumbing to this temptation, he goes on to say, ignores the fact that others make first-person statements and effectively thrusts one into solipsism.

For Shoemaker, as for Wittgenstein and Anscombe, in making certain self-ascriptions one cannot be wrong about whether the subject to which a property is ascribed is oneself. Only self-ascriptions that Wittgenstein considers involving I_s have this peculiar property, and Shoemaker marks them with the somewhat unwieldy terminology as "immune to error through misidentification relative to the first-person pronoun" (Shoemaker 1968, p. 556; hereafter 'IEM'). All others--those Wittgenstein would consider made with I_o as well as those involving self-designating expressions other than 'I'--allow for the possibility that the subject mistakes which object possesses the property in question. He writes:

to say that a statement "a is ϕ " is subject to error through misidentification relative to the term 'a' means that the following is possible: the speaker knows some particular thing to be ϕ , but makes the mistake of asserting "a is ϕ " because, and only because, he mistakenly thinks that the thing he knows to be ϕ is what 'a' refers to.
(1968, p. 557)

Essentially, Shoemaker wants to capture the intuition that certain ways of thinking about oneself simply don't allow for questions to arise about who is doing the thinking. In thinking the thought I would normally express by uttering "I am angry," for example, it seems that I cannot in principle be mistaken about who exactly is angry, for in

feeling angry I cannot find myself in the awkward position of knowing that someone is angry without knowing that it is I who is angry. Feeling angry is inseparable from feeling myself to be angry.

My ascribing anger to myself in virtue of feeling angry is important. For Shoemaker, the distinction between self-ascriptions that are IEM and those that aren't does not consist in the content of the ascription as Wittgenstein and Anscombe believe but rather the way in which the property ascribed is known to be instantiated. Ascriptions of anger to oneself as well as to others employ the same predicate and share a subject-predicate structure. And since the first-person pronoun does refer reflexively, they are true if and only if the tokener is indeed angry. One ascribes anger to others on the basis of observations of their verbal and nonverbal behavior, but one does not normally self-ascribe psychological properties on this basis. Predicates that one self-ascribes without needing to observe one's own behavior "are precisely those the self-ascription of which is immune to error through misidentification" (1968, p. 562).

IEM captures this ascriptive asymmetry without disturbing a commitment to the referentiality of 'I'. Shoemaker writes:

My use of the word 'I' as the subject of my statement is not due to my having identified as myself something of which I know, or believe, or wish to say, that the predicate of my statement applies to it. (1968, p. 558)

The first-person pronoun as it is used in nonaccidental self-ascriptions does refer to oneself, but this reference does not involve self-identification. The relevant ways in which one knows or believes oneself to have a certain property determine whether a self-ascription is IEM, and these ways of knowing are those that do not involve one's identifying oneself at all.

Why think that any self-ascriptions must be IEM? If self-ascribing properties always depended upon identifying something as oneself, Shoemaker (1968, 1986, 1996) argues, we would fall into an infinite regress. He vividly illustrates the point by way of a scene first introduced by the Marx Brothers but now commonly quoted in cartoons, films, and television shows:

Recall the episode in the Marx Brothers movie "Duck Soup" in which Groucho begins to suspect, correctly as it turns out, that instead of seeing himself in a mirror he is seeing, through an empty mirror frame, a double (Harpo, in fact) who is agilely aping his actions. Groucho goes through all sorts of antics in an attempt to fake out and expose the suspected double. Suppose, contrary to the film script, that it really was himself Groucho was seeing in the mirror, and that he became satisfied of this

by seeing that the man in the mirror was performing the very same shenanigans that he himself was performing. Plainly, in order to identify the man in the mirror as himself in this way, Groucho had to know that he himself was performing those movements, i.e., had to know what he could express by saying "I am moving in the ways I see that man moving." To avoid infinite regress, we must allow that at some point Groucho had first-person knowledge that did not rest on an identification. In general, identification-based²⁶ first-person knowledge must be grounded in first-person knowledge that is not identification-based; and the making of introspective judgments is one of the main cases in which this occurs. (1996, p. 211)

To identify something as oneself, one has to know at least some of the properties one possesses in order to determine whether it is oneself. If knowledge of these identifying properties themselves resulted from an identification-dependent process, one would need to have still prior knowledge of those properties used in that process, and so on. Identification must give out, it seems, before one's ability to ascribe properties to oneself.²⁷

²⁶ In more recent work, he has abandoned his 1968 terminology in favor of Evans's coinage, calling the self-ascriptions in question "identification free" (Evans 1982, pp. 179-182; Shoemaker 1994, pp. 196-197). Since Shoemaker's original terminology has become fairly standard in the literature (and since Evans does not consider the same thoughts immune from misidentification), I will use the original phrase to avoid possible confusion. For a recent exploration of the various kinds of immunity, and especially some interesting distinctions between Shoemaker and Evans on IEM, see Pryor (1999).

²⁷ I should note that while this is clearly the main argument he marshals in support of IEM, Shoemaker often writes as though it is

Halting the regress in this way has the consequence--one Shoemaker readily embraces--that IEM self-ascriptions have a kind of priority over all others.²⁸ Thoughts that I would express with I_0 such as "I have a bump on my forehead" are subject to error as Wittgenstein believed, since I could mistake another's bruised forehead in a mirror for mine. In this case, my ascription amounts to asserting that my forehead has a bump, but in saying which forehead is mine I must employ I_s --e.g., my forehead belongs to my body, and my body is the body whose head moves when I move my head (Shoemaker 1968, p. 657). The success of I_s seems to be required even in I_0 cases where errors results. Accordingly, behind one's uses of 'I' as object, and, ultimately, all referential terms, lie self-ascriptions that are IEM. Shoemaker concludes that "[t]here is, I think, an important sense in which each person's system of reference has that person himself as its anchoring point..." (p. 567).²⁹

simply a datum, or at least an entirely separable claim, that certain uses of 'I' are IEM. See, e.g., 1968, p. 561.

²⁸ This type of argument has attracted attention and approbation from philosophers writing in the Continental tradition. Sartre (1968) and more explicitly Henrich (1971) appeal to similar considerations. For a nice discussion of the current and growing points of contact between the two traditions see Zahavi (2000).

²⁹ This consequence brings Shoemaker into close agreement with Chisholm and Lewis, discussed in the previous chapter.

These arguments suggest a particular picture of introspection as well as an explanation for the traditional puzzles that surround it. Hume famously denies that introspection reveals a subject, arguing that that in "gazing inward" he finds only perceptions and no perceiver:

For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe any thing but the perception.
(1978, p. 252)

The word 'observe' in the last clause holds the key to dissolving Hume's puzzlement, argues Shoemaker, since observation is fundamentally tied to identification. But psychological self-ascriptions are IEM, which means that introspection involves no self-identification.³⁰

Still, Hume does seem right to think that in introspecting one's states of mind one does not "stumble" onto oneself as one does often stumble onto others. Shoemaker contends that IEM allows us to explain Hume's insight. Perception of physical objects provides identification information about the objects perceived.

³⁰ Evans (1982, pp. 232-233) makes a similar point in that a subject can know that he stands before a house simply by observing the house. Though the corresponding thought (what he "envisages, or judges") that he would express as "I am in front of a house" amounts to a claim about two spatially related objects, "what he sees does not" (p. 232). We will return to Evans' claim in the next chapter.

Upon seeing my apartment building I learn that it has a limestone façade, for example, and this is the type of information that I can use to pick it out on other occasions. But this is also the kind of information that can lead me to misidentify another building as the one I live in. Similarly, I can glean identification information about myself perceptually, as when I look in the mirror to evaluate the bumpiness of my forehead and discover that my shirt is covered with lint. If introspection were a kind of perception, however, then just as I could be mistaken about whose shirt was linty I could be mistaken about whether it is I who is angry or believe that p. But since I cannot misidentify myself when ascribing psychological properties, introspection must not be a form of perception. We do not need to literally "look inward" via some "inner sense" to ascribe psychological properties to whatever subject we "see" lurking there.³¹

If introspection is not a type of perception, how should we understand it? Shoemaker favors a view in which mental facts are largely "self-intimating" in a weak sense—namely, that there exists "a conceptual, constitutive

³¹ For a nice discussion of such "elusiveness" see the introduction to Cassam (1994) and Chapter 2 of Cassam (1997). See also Kant's similar conclusion that the noumenal self is in principle unknowable (1965). Wittgenstein (1961) offers a more abstruse elusiveness claim.

connection between the existence of certain sorts of mental entities and their introspective accessibility..." (1996, p. 225). For minded creatures like us, introspective access to our mental states is a product of their constitutive causal relations. Our "wiring" is such that when one is in a certain mental state, in the right circumstances a belief will be produced to the effect that one believes that one is in that state. The right circumstances here include possessing a self concept and the relevant mental-state concepts, and one will need to reflect upon one's beliefs. When these conditions obtain, Shoemaker argues that one has highly reliable introspective access to one's mental states and properties.

However reliable our introspective judgments may be, empirical work in psychology (along with the sizeable libraries of testimony offered by many novelists and playwrights) strongly suggests that we do make mistakes about our mental facts. The often-quoted work of Nisbett and Wilson (1977), for example, indicates that subjects do confabulate reasons for their actions according to widely held causal theories about the relations between intention and action. And those susceptible to post-hypnotic suggestion often invent impossible intentions to

rationalize their suggested behavior.³² For this reason, the self-intimation Shoemaker has in mind does not amount to the transparency of the mental often associated with Descartes.³³ If the mind were transparent to itself, all mental states would be conscious, and one could not be in a state without it being known to oneself. One would therefore be infallible and incorrigible regarding one's own states, and few will now accept this "super-certainty."

Still, even weak self-intimation is a controversial claim, and Shoemaker offers an extremely clever argument in support of it. He reasons that if mental facts were not at least weakly self-intimating, then one could believe that *p* while failing to believe that one believes that *p*. Accordingly, it would be conceptually coherent to imagine a "self-blind"³⁴ creature that resembled a blindsight subject's particular lack of self-awareness. After having suffered lesions to the area of the visual cortex called V1, certain subjects deny having any visual experience of objects presented in certain regions of their visual field. Nevertheless, they can correctly answer questions regarding

³² For an extensive and interesting examination of some of the ways in which we fool ourselves, see Wegner (2002).

³³ See, e.g., Meditation II.

³⁴ It's curious that Shoemaker chooses to describe his imaginary subject as "self-blind" given his rejection of a perceptual model of introspection. Perhaps this slip just stands as evidence of the intuitive power of the perceptual model.

what they "saw" presented in their scotoma or blind spot when forced to guess at levels significantly better than chance--hence the term 'blindsight'.³⁵ Similarly, Shoemaker imagines a self-blind creature to be one that has cognitive and conceptual capacities similar to our own but completely lacks first-person access to her own beliefs. As a result, she can self-ascribe properties only on the basis of third-person evidence, including observations of her own behavior or the state of her body.

Does the possibility of a self-blind subject make sense? Could one self-ascribe mental states only in virtue of having characteristically third-person access to one's own states? Shoemaker argues that such a creature is in fact conceptually incoherent. He believes that any creature that possesses our level of rationality and intelligence, along with a conceptual lexicon that equals ours, will "behave in ways that provide the best possible evidence that she is aware of her own beliefs and desires to the same extent that a normal person would be, and so is not self-blind" with regard to intentional states (1996, p. 236). In the case of believing that *p*, both the normal and

³⁵ Weiskrantz (1986) is the locus classicus for the presentation and discussion of this phenomenon. See also Weiskrantz (1997). He attributes the origin of the term to a hasty response to a hasty question regarding the title of a seminar held in 1973 (p. 19).

putatively self-blind person will be disposed to behave similarly. When asked "Do you believe that p ?" rationality dictates that they both respond affirmatively to that question just in case they would answer the question "Is it true that p ?" in the same way (assuming that they both possess the concept of belief). Each will conclude that because p is true, it would be rational to act as if she believes that p is true to maximize the possibility of meeting her goals. Goal maximization also often includes drawing upon the assistance of other believers of p , which in turn means attempting to make others believe that one believes that p . Subjects characteristically encourage others to believe that they believe that p by asserting, among other things, "I believe that p ." Hence, normal and putatively self-blind subjects will be disposed to behave identically in all the relevant ways.

We can put the point another way. Consider, once again, Moore's-paradox sentences. At first blush, a self-blind creature seems like a prime candidate to have and even to express a belief with Moore's-paradox content, since all available evidence could point to p 's being true while the preponderance of her behavioral evidence could indicate her not believing that p . Thus, it seems that she could quite easily find herself in the position of

believing what she would express as "p but I don't believe it."³⁶ But here again rationality dictates that sentences of this form cannot be coherently asserted or assented to for the same reason. Shoemaker argues that one would only assent to "I believe that p" when one would assent to p, and since this is a matter of concepts and their rational deployment, a self-blind creature will be able to recognize their absurdity.³⁷ Hence both the normal and self-blind subject will behave identically in all the relevant ways. And as we saw above, Moore's-paradox sentences only arise when the reporting conjunct is framed in the first-person, present tense. Since a self-blind subject will show the right sensitivity to tense and person on conceptual grounds, it follows that self-blindness is conceptually impossible. Any subject who can self-ascribe mental properties in a third-person way can also do so first-personally simply in virtue of possessing the requisite conceptual and intellectual capacities that we ourselves possess.

³⁶ I owe this way of putting the point to Rosenthal (1995).

³⁷ Though he does think that the (mental) assent conditions for p entail those of one's believing that one believes that p, Shoemaker doesn't think the converse holds (1996, p. 78).

§5. Making Sense of IEM

In Shoemaker's hands, IEM proves to be a powerful explanatory tool. It provides a ready means of distinguishing nonaccidental self-ascriptions from accidental ones in a way that agrees with many of our intuitions about their differences. One's self-ascriptions of psychological properties are immune to errors of misidentification when one is aware of those properties in a particular way. As we might expect, introspection serves as a source of this type of awareness, arising out of the causal relations between one's states, their inputs, and their outputs that are constitutive of one's mental economy. Broadly speaking, according to Shoemaker, for a creature like us to believe that *p* is for her to be aware that she believes that *p*. One's first-order states are quite literally self-intimating--one cannot be in them without being aware that it is oneself who is in them--and hence no room remains for questions regarding whether oneself is the subject of psychological self-ascriptions. Nevertheless, Shoemaker's explanation as to how at least some of one's self-ascriptions involving 'I' are immune to error through misidentification does not conflict with the first-person pronoun functioning as a referring expression. Indeed, if we were not able to refer to ourselves in this

way it appears that reference to anything at all would not be possible.

The neatly unified picture he presents faces its own difficulties, however. Let's reconsider the blindsight-self-blindness analogy as a ready wedge back into Shoemaker's argument for weak self-intimation of mental states. Self-blind subjects begin as imagined analogs to blindsight patients but are eventually ruled out on conceptual grounds. But blindsight patients do not dwell solely in the imaginations of clever philosophers and have been empirically studied for a number of years now. The fascinating feature of such subjects is that they claim not to "see" what is presented in their scotoma, yet they are able to respond correctly to questions regarding that stimuli when forced to guess. Weiskrantz (1997) describes some of the reactions by blindsight subjects to requests for guesses regarding "unseen" objects:

Many subjects, and no doubt experimenters alike, find it embarrassing to pretend that they can guess about something that they cannot see. Indeed, some subjects refuse point blank. I have had a patient tell me, when encouraged to guess in the same way that one would back a horse, "I do not gamble!" Another patient insisted, "my brain is just not set up for guessing!" Another said, "I know it's 50-50 of one or the other. But I can't guess." (p. 66)

This kind of resistance is telling. Blindsight subjects may answer questions correctly when forced to guess what they "saw", but without prompting they have no sense that they have even acquired visual information that can be drawn upon by guessing. As Weiskrantz's anecdotes indicate, the guessing prompted by the examiner seems totally arbitrary to the subject to the point where hazarding a guess at all seems pointless.

How does the position in which a blindsight subject finds herself compare to that of someone who is self-blind? Shoemaker argues that possessing conceptual and intellectual capacities like ours rules out the possibility of a creature who could self-ascribe psychological properties and states in a third-person way but not do so in a first-person way. If asked "Do you believe that p?" a self-blind subject will answer in the same way as she would to the question "Is it true that p?", as would a normal "self-sighted" subject. But what if a self-blind subject is not asked whether she believes that p? Presumably her situation would be quite similar to that of her blindsight analog in that she might not even realize that she has come to believe that p--she wouldn't have a sense as to when or what to ask about her own beliefs even in light of her own

behavioral evidence.³⁸ It seems reasonable to suppose that questions as to whether she believes that p will seem totally arbitrary to her in a way uncharacteristic of what we think of as first-person access.

Of course, Shoemaker could regard the analogy with blindsight as merely a suggestive starting point for his self-blindness argument and not one that runs to any appreciable depth. Nevertheless, conceptual connections by themselves do not guarantee first-person access to one's own mental states. Moore's paradox sentences, for example, wouldn't be absurd to the self-blind subject on purely conceptual grounds. For according to Rosenthal's account favored above, Moore's-paradox sentences are absurd speech acts--they have no coherent performance conditions, but their contents can be true. This means that for a self-blind creature to recognize the absurdity of these sentences, she would have to grasp that the performance conditions for saying "p" and saying "I believe that p" are

³⁸ I should note that Dennett (1991) considers the intriguing possibility that a blindsight subject might learn to use whatever feedback available to eventually become "self-cueing"--so much so that she no longer needs the experiment to force a choice. He goes on to wonder whether such a "self-cueing" subject would then become conscious of the visual stimuli presented in her scotoma. Though he thinks that intuition doesn't favor any outcome, he does think that it's reasonable to suppose that such a subject might eventually become consciously aware of these experiences. Weiskrantz (1997), however, thinks that this possibility is quite remote given that blindsight patients who have been subjects of study have been through thousands of trials with no appreciable change in conscious awareness of visual stimuli.

basically equivalent. But such a grasp does not fall out of possessing the concepts of believing and saying. Rosenthal (1995) argues that our beliefs about the connection between believing and saying participate in a "folk theory so well-entrenched as to constitute commonly shared background knowledge" (p. 327). Accordingly, a self-blind creature's conclusion as to the absurdity of a particular Moore's-paradox sentence "would be mediated by a conscious inference that relies on the relevant folk-theoretic connection" (p. 327). Self-ascriptions made in a characteristically first-person way involve no conscious inferences; hence, however necessary conceptual capacities may be for first-person access to one's mental states, they do not prove sufficient. Self-blindness cannot be ruled out on purely conceptual grounds.

Still, the failure of the self-blindness argument does not directly impact Shoemaker's claim that at least some self-ascriptions must be IEM. His regress argument in support of that conclusion stands independently of an account of self-awareness that results in self-ascriptions that are IEM. Though we risk no conceptual incoherence in imagining a creature that believes that p without believing that she believes that p , it still might be a contingent fact about creatures like us that certain forms of

awareness of mental states preclude the possibility of being mistaken about whether oneself is in those states. Holding onto IEM despite the lack of a conceptual connection between mental states and their reports results in a still weaker self-intimation claim, but one germane to our central question regarding the importance of self-identification to self-ascriptions of various sorts.

As James Pryor (1999) has noted, Shoemaker's characterizations of IEM involve an ambiguity in the notion of misidentification. According to the passage quoted above from his 1968 paper, Shoemaker defines the possibility of error through misidentification relative to a subject term as hinging upon an assumption that a particular object known to be F is identical to the object that a particular subject term refers to.³⁹ The belief that a is F can fall into error relative to the term 'a' in that one could know of some object x that it is F and believe that x is identical to a. For example, upon seeing a woman enter the room in uniform I assert "Marianne is a sergeant." Though I correctly identify the woman's rank, I have mistaken this sergeant for my friend Marianne. Similarly, I could see some particular person reflected in a mirror sporting a broken arm and, mistakenly thinking

³⁹ Cf. Evans (1982), pp. 180-181.

that he is myself, token the belief that I have a broken arm. Pryor calls this type of misidentification "de re misidentification" because it consists in believing a singular proposition about a particular object (1999, p. 274). Roughly speaking, instances of de re misidentification can be described as failed attempts at reidentification.

Though this reading of misidentification conforms to the letter of Shoemaker's 1968 definition, Pryor is right to point out that it doesn't quite capture the Wittgensteinian spirit of the IEM condition clearly in evidence in the Groucho Marx passage above from his 1993 Royce Lectures. The error Shoemaker considers there encourages a more general reading of misidentification, one Pryor terms "which-object-misidentification" or "wh-misidentification" (1999, p. 281). Suppose that I hear a car alarm going off in front of my building. When I reach the street I see several unfamiliar cars with flashing lights and blinking hazards. Looking around, I assert "The black SUV is making all the noise" when in fact the triggered alarm belongs to the station wagon next to it. My error results from my conclusion regarding which car is responsible and not from any beliefs I have about the identity of a specific car. In this case I knew something

to be F but ultimately erred in my conclusion as to which thing is F.

The distinction between de re and wh-misidentification can also account for our treatment of the case of radically mistaken identity raised in the previous section. We described my delusional belief that I am Freud as a case in which I am mistaken about who I am and who is Freud. Pryor would describe this mistake in terms of de re misidentification, for I erroneously believe of some thing (myself) that it is identical with a familiar object (Freud). However, we agreed, following Rosenthal, that even in this extreme case it seems that I can't be wrong about whether I believe that I am Freud. This is to say that my belief is immune to wh-misidentification since it seems that I can't be wrong about which object believes himself to be Freud.

Pryor further argues that wh-misidentification is the more interesting and basic notion, for wh-misidentification entails de re misidentification but not vice versa. If my belief that a is F is immune to wh-misidentification, then I cannot know that something is F while being mistaken about which thing is a. This means that my belief is also immune to de re misidentification, for it doesn't rely upon prior knowledge that a particular object a is F along with

the belief that some familiar object is identical with a. But my belief regarding the noisy car involved no reidentification of a familiar object, and hence it is also immune to de re misidentification. The example was offered as an paradigm of wh-misidentification, however, which means that "immunity to de re misidentification, when based on certain grounds, does not entail immunity to wh-misidentification, when based on those grounds" (Pryor, 1999, p. 286).

Like Shoemaker, Pryor holds that all introspective beliefs about one's mental state are immune to both de re and wh-misidentification,⁴⁰ and both Shoemaker and Pryor describe these beliefs in terms of what one knows about oneself. But even self-ascriptions that are immune to wh-misidentification don't enjoy freedom from error tout court. Shoemaker himself cautions that Wittgenstein's I_o/I_s distinction--and therefore his non-IEM/IEM distinction as well--should not be confused with the corrigible/incorrigible distinction (1968, p. 556).⁴¹ One can apparently be wrong about the property ascribed even when one cannot be wrong about which individual is the

⁴⁰ Pace the remote possibility of what Shoemaker (1970) calls "quasi-memory", which Pryor describes as an example of beliefs that are identification free but still subject to wh-misidentification.

⁴¹ Presumably Shoemaker would also agree that IEM is not tantamount to infallibility.

subject of the ascription. I could hear sounds in the distance and token the thought that I would express as "I hear a woodpecker" when in fact it was a series of gunshots. Though my belief is false, the conditions for immunity to wh-misidentification are satisfied: My auditory experience of hearing the gunshots grounds my belief that someone heard a woodpecker, but those same grounds leave no room for mistakes regarding which subject heard the noise. But if one can indeed be mistaken about which property one instantiates--even if one cannot be mistaken about which person instantiates it--one cannot be properly said to possess knowledge.

Bermúdez (1998) likewise finds talk of knowledge ill-suited to capture the special status of nonaccidental self-ascriptions that Shoemaker intends to seize upon. He writes:

The category of first-person contents being picked out is not defined by its subject matter or by any points of grammar. What demarcates the class of judgments and beliefs that are immune to error through misidentification is the evidence base from which they are derived, or the information on which they are based.... (p. 6)

What matters is the connection between particular evidence bases that serve as sources of self-awareness and the self-ascriptions that those sources support. Not all evidence

bases support self-ascriptions that are IEM, and not all IEM self-ascriptions are true. The question remains as to which evidence bases fund self-ascriptions that can be immune in this particular way.⁴²

We can perhaps make a bit more sense of this connection proponents of IEM have in mind by returning to the account of misidentification introduced in the previous chapter. According to that view, a subject misidentifies herself when she would assent to "a is F" but would deny that she, herself is F when 'a' does in fact refer to her. IEM does settle nicely into this formulation since the denial definitive of misidentification relies upon an component that is ostensibly IEM--namely, that a subject would deny that "I, myself am a" is true in her mouth.⁴³ Following Pryor's lead, a subject wh-misidentifies herself when she would assent to "Something is F" but would deny that "I am F" is true in her mouth apart from any antecedent beliefs she might have about herself. In either

⁴² Pryor does conduct his discussion of immunity to misidentification in terms of grounds that justify a particular belief. However, it's clear that he's interested in characterizing self-knowledge. He suggests, for example, that we have a firm enough grasp of the way in which justification leads to knowledge despite the Gettier Problem (p. 281).

⁴³ This should not be confused with a subject's denying that "I am F" is true in her mouth since she could very well think that she is F while thinking that a is not. The intermediate case is one in which she denies that "I am F" because she would deny the truth of "I am a" in her mouth.

case it seems that self-misidentification can occur only when one already has the capability to self-ascribe psychological properties, and this is just what Shoemaker's regress argument in favor of IEM leads us to expect.

§6. Conclusion

We began this chapter in search of reason to think that nonaccidental self-ascriptions are just those that in principle preclude a subject's misidentifying herself. We found a candidate reason in Shoemaker's argument--inspired by Wittgenstein's remarks on the uses of the first-person pronoun and sharpened by Pryor's observations regarding the ambiguity of misidentification--that at least certain self-ascriptions must be IEM or we face an infinite regress that threatens any explanation of how we are able to self-ascribe properties at all.

As a result, the primary objections raised against NP introduced at the end of Chapter II can now be stated more clearly. Recall that according to NP all self-ascription depends upon self-identification, and self-identification consists in possessing and deploying a battery of information about oneself consisting of a wide range of self-specifying beliefs. One manifests one's ability to identify oneself in the inferences one draws and the

behavior one engages in, including what one is willing to assent to and to deny. Misidentification occurs when a subject falsely believes that some description does or does not apply to her, occasions where her battery contains critical false beliefs or lacks the relevant true ones. This proposal ran headlong into the objections that such a battery of information seemed neither necessary nor sufficient for nonaccidental self-ascription. Intuition suggests that should one suffer even a severe episode of amnesia one could still self-ascribe properties with the first-person pronoun. It also seems intuitively possible for one to possess a rather comprehensive collection of beliefs that are in fact about oneself yet fail to recognize which person that collection picks out.

In effect, a proponent of NP denies that any self-ascription is IEM. But requiring that one's battery of identificatory information fund all of one's self-ascriptions seems to lead to a vicious regress of identification. Accordingly, NP must explain how such a view of self-ascription either involves at most a harmless regress or halts it altogether in a rather banal way acceptable to those who, like myself, do not believe that accidental and nonaccidental self-ascriptions differ in kind.

The persuasive power of regress arguments is directly proportional to the clarity of their starting points, however, for we can settle where they must end only after we understand where and how they begin. Accordingly, to address the regress challenge that Shoemaker raises, in the next chapter we begin with the argument's beginning and end with its ending.

**Chapter IV. The Body and Its Owner:
Immunity and The Self-Ascription of Bodily Properties**

"Would that you dared to believe yourselves--
yourselves and your entrails. Whoever does not
believe himself always lies."

Nietzsche, Thus Spoke Zarathustra¹

§1. Introduction

In Chapter III we pursued clarification of and support for the claim that nonaccidental self-ascriptions do not permit the possibility of self-misidentification. We found that the Wittgensteinian conclusion that certain uses of the first-person pronoun do not pick out their users leads to troubling logical consequences and precludes a satisfactory explanation of Moore's paradox. We then considered Shoemaker's related alternative that certain self-ascriptions are nonaccidental in virtue of being immune to error through misidentification relative to the first-person pronoun (IEM). A subject cannot misidentify herself in nonaccidentally self-ascribing properties, according to Shoemaker, because in making them she does not identify herself at all. He argues that at least some self-ascriptions must be immune to this sort of error, for if all self-ascriptions were subject to errors of misidentification then all self-ascriptions would be

¹ Nietzsche (1954), p. 235.

identification dependent. And if all ascriptions did involve some form of self-identification, it seems that a vicious regress results--one that undermines any explanation of how we are able to ascribe properties to anything at all.

Shoemaker's regress argument presents a formidable challenge to NP. After all, the proposal lies directly in its path: The battery requirement at its heart specifies that all self-ascriptions depend in some way or other upon a collection of self-specifying beliefs, and this requirement seems to precipitate just such a regress of identification. Accordingly, for NP to offer a serious alternative to Shoemaker's Wittgensteinian and Cartesian account of self-ascription, the connection between nonaccidental self-ascription and the battery must be fleshed out.

To better understand what NP must account for, we need a better grip on exactly which self-ascriptions are IEM and therefore nonaccidental. If we follow Shoemaker's argument out, where do we regress to? Shoemaker himself concludes that only certain first-person psychological self-ascriptions based upon introspection could stop the regress. Are all bodily self-ascriptions therefore accidental and open to the possibility of

misidentification? If so, must we settle for a Cartesian-friendly account of nonaccidental self-ascriptions?

These profoundly puzzling questions will occupy and propel us throughout this chapter. I begin with §2 and an attempt to elucidate where Shoemaker's regress argument begins and ends, including what he believes it shows about which self-ascriptions remain vulnerable to self-misidentification and which do not. I do so in part by looking a bit more closely at Groucho's dilemma introduced in the previous chapter. In §§3-4 I consider the possibility that the regress may end elsewhere by examining arguments given by Gareth Evans and José Luis Bermúdez respectively that certain bodily self-ascriptions can also be IEM, but they prove unsustainable. I devote §5 to a discussion of some empirical work suggesting that bodily self-ascriptions--even those made on the basis of normally functioning senses--are indeed subject to errors of misidentification. In §6, I argue that NP can accommodate the empirical data and explain our sense of body ownership in terms of one's body image that arises out of one's battery of self-specifying beliefs. Finally, this explanation provides shape and direction for an explanation of nonaccidental psychological self-ascriptions undertaken in the next and final chapter: The self-ascription of

bodily properties at best only seems nonaccidental, but errors of ownership cannot be ruled out. How things seem rarely settles how things are, but NP must speak to both.

§2. Understanding the Regress: Beginnings and Endings

We can view Shoemaker's regress argument as having both positive and negative implications. If sound, its negative impact is fairly obvious: Identification must come to an end before one's ability to self-ascribe properties; hence, accounts like NP that explicate all instances of self-ascription in terms of one's ability to identify oneself attempt the impossible.

The positive implications are a little less clear, however. Even those who find the argument compelling--and who agree that at least some self-ascriptions must be IEM--do not believe that the motivation for launching the regress clearly determines where it must end. Differences arise as to whether only psychological self-ascriptions can be IEM or whether certain self-ascriptions of bodily properties must also preclude self-misidentification.

To help sort out this dispute, we first need to understand the regress itself. Broadly speaking, if an ascription of the form "a is F" depends upon identification, it can be regimented as a conjunction of

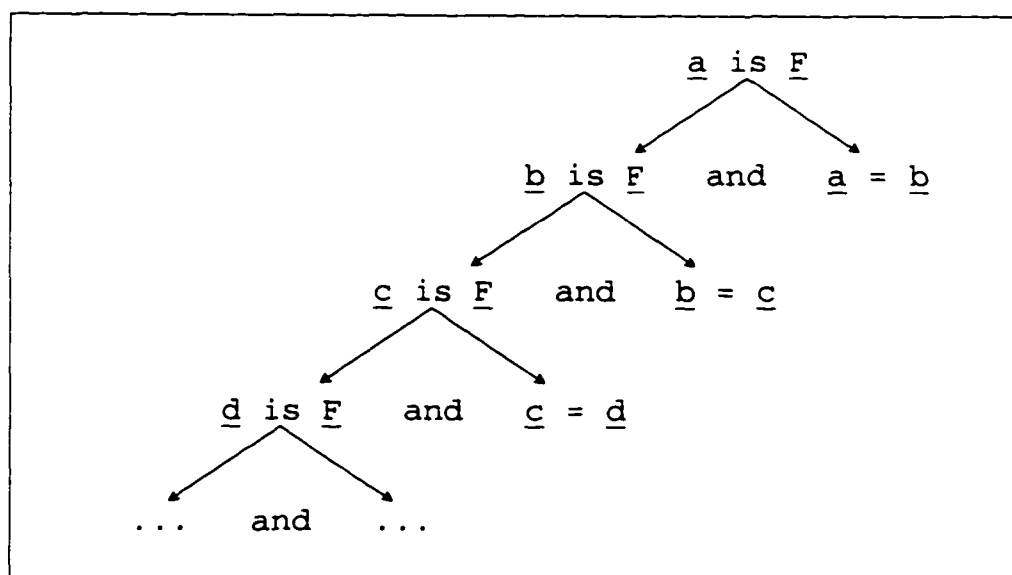


Figure 1. Regress of Identification

the beliefs "b is F" and "a = b."² If the ascription of property F to presented individual b is itself identification dependent, it too can be regimented as the conjunction of "c is F" and "b = c." And so on, as in Figure 1 above.

In its simplicity, however, the above picture emphasizes the generative role of the property ascription and underplays the role of the identificatory beliefs in perpetuating the regress. The identification component ("a = b") and the identification inference in which it figures are in fact to blame. An individual a is familiar to one because one possesses a battery of a-specifying beliefs {a is G₁, a is G₂, ..., a is G_n}. Accordingly, to ascribe the

² The identity sign should be read as numerical ("identical with") and not qualitative identity ("identical to").

property F to a familiar individual a in a way that involves identifying it, one must believe that a has some set of properties $\{G_1, \dots, G_n\}$. One must also believe that some object b instantiates F as well as $\{G_1, \dots, G_n\}$ (or some significant proper subset thereof). One identifies b as a in virtue of the overlap between a- and b-specifying beliefs, and since one believes both that b is F and that a is identical with b, one concludes that a is F.³ To explain the ascription "b is F" in terms of an identification of b with another familiar individual c, it seems that one would also need a battery of c-specifying beliefs to identify b with c. No end of identification seems to be in sight. Batteries of belief make identification possible, but they apparently must be acquired in some way that itself does not involve identification.

We can trigger the regress in a more concrete way by first asking how one comes to believe that some individual other than oneself is F. Suppose that I meet Vivienne at Grand Central Terminal and discover that she has dyed her hair red. How do I do this? I don't merely believe that some individual or other has red hair, I must pick Vivienne out from the crowd. If she is already familiar to me, I

³ This chain of inferences need not be conscious.

have a wealth of useful information to draw upon to do so. I immediately look for women of a certain hair color, height, dress, age, etc., and I pay particular attention to women standing in specific locations in Grand Central looking like they might be waiting for someone.⁴ My beliefs needn't be exhaustive or entirely accurate--indeed in this case one of my identifying beliefs about her is false. Still, I believe that Vivienne has certain properties, and I look for some individual who possesses most if not all of those properties.⁵ In scanning the flow of people, I see someone who resembles Vivienne in many ways apart from hair color, and I conclude that the red-haired woman is Vivienne. If I'm correct, we head off for lunch. Of course, my antecedent beliefs about Vivienne may lead me to mistake similarly featured individuals for her; with identification comes the possibility of misidentification. Nevertheless, it goes almost without

⁴ Face recognition is arguably a different story. The wonderful and extensive work on this capacity strongly suggests that we have dedicated and redundant subpersonal systems devoted to this important task. See the papers in Young (1998) for representative work. Still, we can make a distinction between recognizing a set of familiar facial features (identifying a face as Vivienne's, say) and observing the features of a face seen for the first time.

⁵ Our identifications of those familiar to us are dynamic, of course. I could discover upon meeting Vivienne that she has dyed her hair. If I were to meet her again next week I would have different beliefs about her hair color to aid in picking her out from the crowd.

saying that if I did not have at least some notion of who I was looking for I could not succeed in recognizing anyone.

Does it make sense for one to ascribe a property to oneself in this manner? We can imagine cases in which it does. Consider once again Groucho's dilemma. Recall that in "Duck Soup" he finds himself faced with ambiguous perceptual cues suggesting that what he sees before him may be a reflection of himself or may be another person imitating his actions and dress. He attempts to resolve the dilemma by performing experiments that involve acting in particular ways and noting the reactions of his conspecific. In the film, Groucho eventually does succeed in determining that his putative "reflection" is in fact another (Harpo, as it turns out) attempting to fool him. Though amazingly adept at aping Groucho, Harpo fails to mimic his movements completely--Groucho notices the incongruities, and Harpo's ruse is discovered.⁶

What incongruities does Groucho notice? Shoemaker claims that Groucho's recognition (either that he is in front of a mirror or that he is in front of an imitator) depends upon his ability to self-ascribe certain

⁶ To be more precise, the famous "mirror" scene ends--to great comic effect--when Groucho sees Chico (also dressed like Groucho) stumble into Harpo. Perhaps Harpo could have successfully fooled Groucho into thinking that the mirror had not been broken after all.

psychological states and properties, including what he would express as "I am moving in the ways I see that man moving" (1996, p. 211). It's important for Shoemaker that this belief involve the first-person pronoun or its mental analogue. Its extensional equivalent, in this case, "Groucho is moving in the ways Groucho sees that man moving" leaves open the possibility that Groucho fails to believe that he, himself is Groucho. Perhaps he can notice that the movements of the unidentified individual in front of him do not completely match his own without harboring many beliefs about who he is, but it seems that Groucho can't be mistaken as to whether he, himself is moving in such-and-such a way.

Could Groucho's beliefs about his movements themselves depend upon his recognizing some unidentified moving thing as himself?⁷ He must have some basis for wondering whether what he sees moving is indeed himself, some set of properties he believes true of both he, himself and the unidentified individual that could support an identifying inference. In the case at hand these properties are largely observational: Groucho sees that they both sport similar glasses, eyebrows, and mustaches,

⁷ I'm assuming for the moment that no question arises for Groucho as to how mirrors work.

and that they wear identical nightgowns and caps. And the evidence regarding Harpo's movements is also observational.

Could he find himself in a similar situation in which he is presented with an unidentified individual that is in fact himself but does not involve a mirror? Here again, it seems that he might. He might not be sure, for example, that the limb he sees moving is in fact his own. He might even wonder whether the body he experiences in certain ways belongs to him.

How could he extricate himself from such a peculiar predicament and come to believe what he would express as "This body is my body" or "This arm is my arm"?⁸ It seems that he must believe that some presented individual is $\{G_1, \dots, G_n\}$ and that he, himself shares many or most of those properties. In the odd case we're imagining, these properties will include psychological self-ascriptions such as "I'm intending to move in this way" that may then overlap with observations about bodily movements.⁹ For one to identify a particular body as one's own one must already

⁸ He need not pick out his body demonstratively. He could do so descriptively--e.g., "the body wearing the nightgown and cap currently moving ... is mine." I defer to the demonstrative case for the sake of brevity.

⁹ Interestingly (an anticipating somewhat), we can frame the traditional interactionism problem for the dualist in a different way. We see that the dualist also faces an identification problem as well--that is, how one could identify a body as (contingently) one's own in virtue of possessing information about completely unrelated properties.

have certain beliefs about one's psychological states and properties that one can then use to compare with candidate behavioral consequences of those states and properties.

What about the next step? Could Groucho be called upon to recognize a presented unidentified self with certain psychological properties as himself? Shoemaker argues that if he could we have no hope of avoiding vicious regress. For Groucho to come to believe that he is, e.g., angry on the basis of first observing that some unidentified self is angry and then concluding that he is identical with that self, it seems that he must already possess some identifying beliefs about which properties he currently possesses. If those identifying beliefs themselves resulted from identifying some self that possesses them as oneself, one would need still further identifying beliefs to identify oneself in that way. And so on.¹⁰ Hence, Shoemaker concludes that the regress must end with certain psychological self-ascriptions that do not rely upon any self-identification whatsoever--and hence are IEM--or it must not end at all.

¹⁰ Shoemaker gives versions of this argument in 1968, 1986, 1996. He frames the argument in terms of self-knowledge instead of self-belief, but in light of the previous chapter, the issue is independent of the epistemic statuses of identifying beliefs.

Indeed, self-misidentification seems to require that one self-ascribe properties that are IEM. Intuitively, any conclusion that Groucho draws regarding the identity of the "image" before him hinges upon his already possessing certain beliefs about himself. We can imagine him expressing the incorrect conclusion (in this case) that the "image" he sees before him is in fact his own as "I see myself in the mirror" or "That's me." Roughly speaking, it appears that one must take certain beliefs about oneself for granted before one can identify or misidentify oneself in various ways. And Shoemaker gives us a way of understanding this appearance in terms of certain beliefs about one's psychological states and properties that can anchor an identification of a presented individual as oneself.

According to Shoemaker, then, Wittgenstein was indeed correct to hold that not all first-person self-ascriptions are unique with regard to identification. Since Shoemaker thinks that one's body is only contingently one's own,¹¹ the self-ascription of bodily properties is subject to misidentification. Those such as "I am raising my arm" rely upon other self-ascriptions such as "my body's arm goes up when I raise it." Self-identification at the

¹¹ He says as much in his TLS review of Cassam (1997)

psychological level cannot occur without pain of regress, however. Accordingly, it makes no sense to ask oneself "Someone is raising her arm, but is it I who is raising her arm?" when one is conscious of someone's arm being raised, for then one is conscious of oneself as raising one's arm. Shoemaker concludes that awareness of oneself as an object can play no role

in explaining my introspective knowledge that I am hungry, angry, or alarmed. This comes out in the fact that there is no possibility here of a misidentification; if I have my usual access to my hunger, there is no room for the thought "Someone is hungry all right, but is it me [sic]?" (1996, p. 211).

It's important to note just what kind of question Shoemaker thinks is being asked. He argues that the regress must end with self-ascriptions based upon introspective awareness, and so the question should be filled out as "Someone is introspecting that he is hungry, but is it I?" Hence, the question does not seem to be a who question--one that can be answered by appeal to the battery of commonsense self-specifying beliefs as NP would have it. Instead, the question asks whether one can be mistaken as to whether oneself, whoever one is, is introspecting a particular mental state or property. If

that question always makes sense, regress appears unavoidable.

§3. Evans: Proprioception, Bodily Self-Ascriptions, and IEM

Shoemaker takes the regress argument to show that at least some psychological self-ascriptions must be IEM.

Accordingly, questions of the form "Someone is F, but is it I who is F" prove nonsensical only for certain psychological states and properties that one is aware of via introspection. Since it yields self-ascriptions that are IEM, introspection cannot be a form of perceptual awareness or awareness of oneself as an object. As a result, questions of body ownership can arise, and hence no self-ascriptions of bodily properties can be IEM.

Gareth Evans (1982) also believes that self-identification must give out before one's ability to self-ascribe certain properties on pain of regress. He agrees with Shoemaker that if one always self-ascribed properties with thoughts and assertions of the form the Φ is F for any self-designating description Φ , we would fall into regress, one that precludes any explanation of how one comes to believe that oneself is F. A gap always remains, Evans argues, between grasping that the Φ is F and grasping that I am F, for no description Φ guarantees that in grasping

that the Φ is F one thereby realizes what one would express as "I am F" (1982, pp. 255-256). Hence, if one explained one's grasp that the Φ is F is a self-ascription in virtue of another description Ψ --that the Ψ is identical with the Φ , e.g.--one would then need to explain one's grasp that Ψ is self-designating or describing. And so on.

Unlike Shoemaker, however, Evans bristles at the idea of explaining the peculiar features of nonaccidental self-ascriptions in Cartesian-friendly terms such as mental state self-intimation. If we do conclude that the class of nonaccidental self-ascriptions is exhausted by those based upon introspective self-awareness, we risk succumbing to the impression that "in thinking of oneself self-consciously, one is paradigmatically thinking about oneself as the bearer of mental properties, or as a mind--so that our 'I'-thoughts leave it open, as a possibility, that we are perhaps nothing but a mind" (p. 217, emphasis his).

Consequently, Evans contends that many nonaccidental self-ascriptions rest upon certain ways one has of gaining information about oneself as a physical thing located within the physical world. Self-ascriptions funded by information regarding the state of one's body and one's position in the world--when received in the normal way--are IEM. For, according to Evans, the theoretical necessity of

IEM self-ascriptions demonstrates that particular sorts of information one receives about oneself impact one's self-ascriptions "directly and immediately" (1982, p. 224). That is, self-ascriptions are IEM because self-conscious thought demands that a subject "be disposed to have such thinking controlled by information which may become available to him in each of the relevant ways" (p. 216). And one's nonaccidental self-ascriptions are also sensitive to information about one's body when gained in particular ways.¹² Questions of the form "Someone has his legs crossed, but is it I?" will likewise be nonsensical in certain cases.

Evans believes that two types of information can give rise to bodily self-ascriptions that are IEM. The first he describes as "a general capacity to perceive our own bodies" from, as it were, the inside (1982, p. 220). This capacity consists of various monitoring systems such as somatic proprioception, one's sense of balance, kinesthetic feedback, etc. Such systems provide a stream of information regarding the state of one's body, the position

¹² In direct contrast to Shoemaker, Evans holds that nonaccidental self-ascriptions are manifested in action (like the information Groucho used to recognize his mirror reflection), but not in terms of getting an object (one's body) to act. In his words, "I do not move myself; I myself move" (1982, p. 207).

of limbs, skin and joint tension, vestibular feedback during motion, etc.¹³ Evans writes:

There just does not appear to be a gap between the subject's having information (or appearing to have information), in the appropriate way, that the property of being F is instantiated, and his having information (or appearing to have information) that he is F; for him to have, or to appear to have, the information that the property is instantiated just is for it to appear to him that he is F. (p. 221)¹⁴

To receive proprioceptive information regarding limb position is to receive information regarding the position of one's own limbs. Hence, asking oneself "Someone's arm is raised, but is it I" does not make sense when the first clause expresses information gained about one's body through proprioception.

This claim seems particularly vulnerable to innumerable counterexamples. We could imagine, for example, that another's arm is linked to one's somatosensory system in such a way that one receives feedback "from the inside" about its movement and position.

¹³ For a more extensive list of the informational systems that constitute somatic proprioception, see the general introduction to Bermúdez, Marcel, and Eilan (1995), pp. 13-15. There they also distinguish between proprioceptive information and proprioceptive awareness. They characterize informational states as subpersonal and only states of awareness as representational in the way that interests us here. Accordingly, in talking of proprioception I mean representational states resulting from processing by the proprioceptive system.

¹⁴ Here Evans is using a bit of shorthand. When saying that a subject has information that "he is F" he means "he, himself is F."

Isn't this a relatively compelling case in which it would make sense to ask "Someone is raising his arm, but is it I?"

Evans anticipates this type of objection, and he believes that he can disarm it. Like Shoemaker, Evans frames IEM in terms of what one knows and the presuppositions for that knowledge. He claims that a subject who self-ascribes what she would express as "I am raising my arm" on the basis of information gained via a deviant somatosensory causal chain does not in fact possess knowledge about the position of any arm. IEM self-ascriptions of bodily properties rely upon the presupposition that the relevant information results from the normal functioning of one's senses. If that presupposition is violated, a subject who believes that her arm is raised has a false belief and hence does not know what she self-ascribes.

Interestingly, Evans does agree that error is possible. The question "Someone's arm is raised, but is it I?" does make sense if the subject is aware that she is connected to another's arm in a deviant way. The prevailing assumption in standard cases that the senses are functioning normally can indeed be called into question. And if one is told that one's senses are not functioning

normally, one cannot trust the information gained in that way. But even in deviant cases we imagine that one receives feedback, for example, as if one's own arm is raising. It does not show, argues Evans, that in the normal case one assumes that the subject reported on by proprioception is identical with oneself. It at most shows that finding oneself in the position of questioning whether in proprioceiving an arm raising it is one's own is to assume that one's senses are not functioning normally.

The second way of gaining information about oneself that funds bodily self-ascriptions that are IEM, according to Evans, "has an importance in our thought about ourselves which it is difficult to exaggerate" (p. 222). We seem able to locate ourselves in the world simply in virtue of perceiving the objects around us without also perceiving ourselves. To take one of Evans's examples, I can come to believe that I am in front of a house by merely perceiving a house (p. 232). Nothing more "on the side of awareness" is called for (p. 231).

Though apparently "we only have to be aware of some state of the world in order to be in a position to make an assertion about ourselves" (p. 231), Evans cautions that we should not be misled, as Hume was, by this fact. My self-ascription that I am in front of a house based upon this

form of perceptual awareness brings to bear my concept of
 ... IS IN FRONT OF A HOUSE. To possess that concept, my
 abilities to apply it must accord with what Evans calls the
 "Generality Constraint":

If we hold that the subject's understanding
 of 'Fa' and his understanding of 'Gb' are
 structured, we are committed to the view
 that the subject will also be able to
 understand the sentences 'Fb' and 'Ga'. And
 we are committed, in addition, to holding
 that there is a common explanation for the
 subject's understanding of 'Fa' and 'Ga',
 and a common explanation for his
 understanding of 'Fa' and 'Fb'. (1982, p.
 101)

The Generality Constraint imposes requirements of
 substitutivity on both the subject concept ('Idea' in
 Evans's terminology) and the predicate concept. The
 cognitive abilities required to token states with
 conceptual content themselves enable the tokener to
 decompose and to generate new contents based upon her
 conceptual lexicon. Generality dictates, for example, that
 if one can entertain the conceptualized thoughts that
 elephants are large and that airplanes are shiny, one will
ipso facto be able to think that airplanes are large and
 that elephants are shiny. One will also be equipped to

draw related generalizations and inferences regardless of the proximity of elephants and airplanes.¹⁵

The applicability of the Generality Constraint to first-person self-ascriptions is rather straightforward. To possess a concept of a particular property such as ... IS IN FRONT OF A HOUSE, I must be able to apply that concept to a host of objects and individuals other than myself. If I also possess the concept of CAR, I can token the thought that I would express as "The car is in front of the house." To grasp the concept is to understand that it applies to physical things spatially located within the world, and to apply it to oneself is to grasp that oneself is also a substantial and persisting self. Hence, Evans writes:

No judgment will have the content of a psychological self-ascription, unless the judger can be regarded as ascribing to himself a property which he can conceive as being satisfied by a being not necessarily himself--a state of affairs which he will have to conceive as involving a persisting subject of experience. (p. 232)

The notion of an ascribing, perceiving subject is bound up with the idea of its being located within an objective, spatially ordered world.¹⁶

¹⁵ See, e.g., Hurley (1998).

¹⁶ Evans is here taking up Strawson's concerns regarding singular thought and the preconditions for thinking about objects.

Since one's thoughts about oneself are also dispositionally sensitive to information of this type, Evans argues that self-ascriptions based on this information are likewise IEM. As in the case of a deviant proprioceptive link, Evans contends that one cannot know the spatial location and orientation of an object when known about in this way without knowing which object one has knowledge of. If one was receiving information about the world from the normal operation of one's senses, then, for example, the question "Someone is in front of a house, but is it I?" is nonsensical. And if one has reason to think that one's senses are not functioning normally--that one's visual experiences are the result of powerful hallucinogens, e.g.--then error may be possible. But this again shows only that one has switched assumptions and not that one needs to identify some physical object located in front of a house as oneself in the normal case.

Moreover, if one did need to identify which object was in front of a house when one perceives only the house, Evans contends that it's no longer at all clear how one could ever make sense of one's being located somewhere in the world. His remarks as to why he thinks this are

incomplete,¹⁷ but perhaps we can make sense of his point as a regress worry similar to Shoemaker's. If perceptual awareness cannot provide information that settles the question of self-location and self-identification, what could? To bridge the gap between thinking that the thing located in such-and-such a place is F and thinking that I am F, I would need to identify myself as the thing located in such-and-such a place. To do that I would need to believe that I had properties in common with some particular physical thing located in a spatially extended world in virtue of which I could conclude that I am that thing. The only relevant properties I could share with some spatially located individual are spatially locating ones. Hence, it seems that if a subject had no capacity to self-ascribe properties that placed her within an objective spatial order in a way that was IEM, she could never locate herself within the world.¹⁸

¹⁷ John McDowell, the editor of Evans's Varieties of Reference, gestures towards Evans's reasoning for this conclusion in the appendix to Chapter 7, but it's obvious from his remarks that Evans had not completely worked through this issue.

¹⁸ Evans might also frame this argument in terms of the concepts one must be able to grasp and to apply to oneself. That is, to grasp the concept of ... IS IN FRONT OF A HOUSE I must be capable of applying it to objects possessing physical and spatial properties. To extend the applicability of such a concept to myself, I would have to grasp that I am a physical person like any other. But that is what grasping the concept was supposed to provide.

Overall, then, the types of information that a subject's self-ascriptions are sensitive to combined with her conceptual repertoire require that at least certain self-ascriptions of bodily properties be IEM. As a result, Evans concludes that "[o]ur thoughts about ourselves are in no way hospitable to Cartesianism" since they involve both mental and physical aspects of oneself with equal intimacy, immediacy, and cognitive importance (p. 256). If every way of gaining information about oneself depended upon identification, no explanation of first-person self-ascription, including self-locating beliefs, could be given.

Though Evans's view provides us with a powerful way to resist Cartesian temptations that can easily influence explanations of nonaccidental self-ascriptions, it is not without deep problems. First, Evans contends that self-specifying information derived from both proprioception and perception do not involve an identification assumption but rather an assumption that one's senses are functioning normally. The obvious difficulty here is the threat of circularity. If by normal functioning Evans means that when working in that way the senses give rise to self-ascriptions that are IEM, he cannot, of course, explain IEM in terms of the normal functioning of the senses.

Perhaps normal functioning and IEM need not be interdefined. We could instead understand "normal" as the senses functioning in a way that accords with their evolutionary design.¹⁹ This response does disarm any immediate circularity in his construal of normal function. But avoiding circularity in this way simply passes it on. Evans would still need to motivate in a non-question-begging way the claim that the evolutionarily selected function of the senses is to provide self-conscious self-specifying information as opposed to information that in fact happens to be about oneself.

In any event, notice that even the deviant cases that Evans considers are not necessarily those in which one's senses fail to function normally. The abnormality lies not in the sensing but in the range of things sensed. Indeed for one to proprioceive the raising of another's arm one's proprioceptive sense must be functioning normally. (We shall return to this point below.) And as we saw in the previous chapter, Shoemaker holds that a false self-ascription can be IEM, and this makes intuitive sense. My false belief that I would express as "I hear a woodpecker in the distance" seems no less direct and immediate "from the inside" than if I had actually heard a woodpecker

¹⁹ I thank Michael Levin for pressing this point.

instead of gunshots. Given his strong reliance upon epistemological conditions, however, Evans cannot agree. Hence, claiming that one cannot possess knowledge of any individual when receiving sensory information in abnormal conditions risks misplacing the phenomenon.²⁰

Perhaps more importantly, Evans characterizes nonaccidental self-ascriptions in terms of their dispositional sensitivity to certain forms of information about oneself. However, a dispositional account of self-ascription seems ill-suited to explain occurrent thoughts about oneself. For if a subject lacked the relevant sort of information, she should be unable to ascribe properties to herself in the relevant way.

Consider the following thought experiment given by Anscombe (1975):

And now imagine that I get into a state of 'sensory deprivation'. Sight is cut off, and I am locally anaesthetized everywhere, perhaps floating in a tank of tepid water; I am unable to speak, or to touch any part of my body with another. Now I tell myself, "I won't let this happen again!" If the object I meant by 'I' is this body, this human being, then in these circumstances it won't be present to my senses; and how else can it be 'present to' me? But have I lost what I mean by 'I'? Is that not present to me? Am I reduced to, as it were, 'referring in

²⁰ Pryor (1999) makes a similar point regarding Evans's treatment of quasi-memory. That is, his claim that one does not have knowledge in such cases seems unmotivated.

absence'? I have not lost my 'self-consciousness'; nor can what I mean by 'I' be an object no longer present to me. (p. 152)

Though Evans agrees with Anscombe that a sensory-deprived subject would still be able to self-ascribe properties nonaccidentally, he believes that a subject need not have any occurrent information about herself to know which object she is dispositionally related to. It is enough that the subject is still dispositionally related to one object and one object alone. Her thoughts about herself would indeed be controlled by sensory and bodily awareness were there any.²¹

In Anscombe's case, however, one's senses are ex hypothesi functioning normally, but one receives no sensory information at all. Far from being a outlandish philosophical fantasy, sensory deprivation tanks are now available for home purchase.²² If one's limbs are not allowed to touch, it is reasonable to assume that one will not be able to specify their location. Assuming also that one is well-fed, not suffering internal injury before

²¹ Indeed, Evans explicitly offers (p. 216 n21) that characterizing one's ability to think about oneself in terms of the dispositional sensitivity of one's nonaccidental self-ascriptions to sensory information circumvents Anscombe's argument.

²² A simple search of the World Wide Web will fetch a host of links to manufacturers. My favorite is the cleverly named "Think Tank International," available on line at <http://www.thinktank.com.sg/>.

stepping into the tank, etc., somatic proprioception will be of little use in locating oneself. And if one can nonaccidentally self-ascribe properties without being able to locate oneself at a particular moment, no source of sensory information is necessary for one to think about oneself in the relevant way.

Hence, it's not clear how merely possessing a disposition to have one's thinking controlled by the relevant information can bear the necessary philosophical weight. After all, how can a dormant disposition serve to ground occurrent first-person thought? How can we be assured that one has not lost track of the physical object that is in fact oneself? Hence, if one cannot succeed in locating oneself spatially yet one can still unproblematically think the thought that one would express as "I won't let this happen again," then a dispositional view like Evans's cannot be correct.²³

Evans risks begging another question.²⁴ He claims that our conception of ourselves as spatially located

²³ O'Brien 1995 has a detailed and fairly clear statement of this objection to Evans.

²⁴ Jerry Fodor (2003) makes a similar claim of circularity in explaining concept possession in terms of what one knows about what falls under its extension or what inferences it participates in. A related circularity objection haunts what Evans calls Russell's Principle, or the principle that for one to think of an object one must have a "discriminating conception of that object--a conception which would enable the subject to distinguish that object from all other

persisting selves falls out of the fact that our thoughts about ourselves conform to the Generality Constraint. But it's difficult to see how merely grasping the concept of ... IS IN FRONT OF A HOUSE and its combinability with the subject concepts one possesses entails that one can properly apply it to oneself. One would need the further belief that one is the kind of thing that can be in front of a house--viz., a physical thing that can be spatially located. If a subject must know what falls under a concept's extension to possess it, then Generality ceases to be explanatory. For if possessing a concept with implications for self-location demands that one know what falls under its extension, then one would already have to know that one is a persisting, spatially located thing before one could apply it to oneself in a way that was not arbitrary.

§4. Bermúdez: Nonconceptual Bodily Self-Ascriptions and IEM

Evans's attempt to argue that certain self-ascriptions of bodily properties must also be IEM proves well motivated in its flight from Cartesianism, but it also proves decisively

things" (p. 65). Whereas the Generality Constraint enjoys fairly wide acceptance, many have taken issue with Russell's Principle for reasons irrelevant to our present concerns. One could perhaps loosen that restriction on self-thought while remaining faithful to Evans's general position. Bermúdez's approach to be discussed later could be viewed in this way.

fraught with difficulties. But he is not alone in thinking that certain self-ascriptions of bodily properties are like certain psychological self-ascriptions in the types of error they are subject to and immune from.

In The Paradox of Self-Consciousness, José Luis Bermúdez follows Evans in pursuing the conclusion that certain bodily self-ascriptions are IEM. His strategy differs from that adopted by Evans, however, in that he tries to motivate the theoretical necessity of states with nonconceptual content that are IEM and therefore count as nonaccidental self-ascriptions of a sort. This approach has many benefits. It avoids, at least prima facie, begging questions concerning concept possession. In addition, Bermúdez does not give a dispositional account of nonaccidental self-ascription, which also avoids objections raised against such views in the previous section. Moreover, extending the range of types or forms of content that can be correctly characterized as genuinely nonaccidental offers a way of dispelling the mystery of how human psychology eventually comes to be shot through with the richer, conceptual forms of self-consciousness exemplified in the nonaccidental self-ascriptions of various properties.

For this strategy to succeed, however, Bermúdez must provide plausible criteria that a state with nonconceptual content must satisfy to be considered nonaccidental. Thus far we have focused on various attempts to explain the uniqueness of nonaccidental self-ascriptions in terms of their immunity to a certain sort of error. Accordingly, if we hold both conceptual and nonconceptual states to the same theoretical standards, the challenge lies in extending the applicability of IEM to states with nonconceptual content where talk of error is not yet clear. We now turn to Bermúdez's attempt to do just that.

Elucidating exactly what nonconceptual content in general amounts to is a difficult task. Bermúdez himself believes that establishing the existence of states with autonomous nonconceptual content disarms the circularity in a certain explanation of nonaccidental first-person thought, one he terms the paradox of self-consciousness.²⁵ This paradox roughly amounts to the circularity that results from analyzing self-consciousness solely in terms of a subject's mastery of the first-person pronoun. Such mastery requires that a subject know--or at least believe--

²⁵ By self-consciousness Bermúdez means possessing states with nonaccidental first-person content, or content that is both IEM and has immediate implications for action. I will use 'self-consciousness' and 'nonaccidental first-person thought', and 'nonaccidental self-ascriptions' interchangeably when discussing his work.

that she, herself is the token of the speech act. But spelling out what it is for a subject to believe that "she, herself" is such-and-such involves reference to the type of thought that we want to explain.

It's not clear, however, that Bermúdez's concern achieves the status of paradox. And circularity doesn't seem inevitable. Perhaps a better way of casting the problem--as we have done throughout--is in terms of how nonaccidental self-ascriptions, whether in thought or in language, differ from their accidental correlates. This in effect denies that we have two disparate things--first-person thoughts and their expressions--that can be viciously interdefined.²⁶

Nevertheless, attributing nonconceptual representational states to creatures lacking linguistic abilities makes sound theoretical sense. Certain intentional systems--including non-linguistic animals and pre-linguistic infants--arguably lack concepts. Still, Bermúdez contends that plausible explanations of the behavior of any intentional system require ascribing representational states to it because no law-like relation holds between its sensory input and behavioral output.

²⁶ Bermúdez no doubt formulates his problem in the way he does because of his beliefs about the possession conditions for concepts. We will come to this a bit further below.

Differences in behavior in the face of the same sensory input indicate that a creature's behavior is a function of a complex group of states, some of which differ from those present on a previous occasion (a past predator can become prey, e.g.), or perhaps misrepresent a current state of the world or its body.

Proponents of nonconceptual content also like to draw attention to the richness of perceptual experience.²⁷ One seems able to represent more properties than one can describe, and this suggests that the range of one's representational states exceeds one's conceptual repertoire.²⁸ I can discriminate two subtly differing shades of green when seeing them side by side yet fail to distinguish them when they are presented in succession. And my thoughts about either shade are severely limited in their absence. My perceptual experiences represent the two colors, but those representations cannot support inferences or participate in contents in the absence of the properties represented.

Broadly speaking, any state that supports generality will be conceptual. Accordingly, states with nonconceptual

²⁷ See, e.g., Peacocke's argument that "an experience is not restricted in its range of possible contents to those points or ranges picked out by concepts ... possessed by the perceiver" (1992, p. 68).

²⁸ This is not universally accepted, however. See McDowell (1994).

content, though representational, must then be holophrastic, lacking the necessary structural features to make them productive and systematic in a way that respects the Generality Constraint. Could such states qualify as primitive forms of self-consciousness? Bermúdez suggests that to be considered forms of primitive self-consciousness, they must meet the requirements for any representational state to be self-conscious--namely, they must have immediate implications for action, and they must be nonaccidentally about oneself (1998, pp. 147-148). Skipping the former for present purposes, thoughts are nonaccidentally about oneself, Bermúdez and many others argue, because they are IEM.

Like Evans, Bermúdez characterizes IEM in terms of the various ways creatures have of obtaining information about themselves. Unlike Evans, however, he agrees that talk of knowledge per se is ill-suited to capture the special status of nonaccidental self-ascriptions that Shoemaker intends to regiment. He writes:

The category of first-person contents being picked out is not defined by its subject matter or by any points of grammar. What demarcates the class of judgments and beliefs that are immune to error through misidentification is the evidence base from which they are derived, or the information on which they are based.... (p. 6)

What matters is the connection between particular evidence bases that serve as sources of self-awareness and the representational states that those sources support. Not all evidence bases support states that are IEM, and not all contents that are so immune are true.

Bermúdez offers a fairly straightforward argument in favor of nonconceptual contents that are IEM resulting from information obtained about the state of one's body via certain sensory channels. Once again, consider states resulting from somatic proprioception. These are representational states because they, like any other representational state, are causally sensitive to sensory input and mediate behavioral output. Moreover, somatic proprioception does seem to provide the right kind of evidence base, for it apparently "cannot give rise to thoughts that are accidentally about oneself" (1998, p. 148). Bermúdez continues:

One of the distinctive features of somatic proprioception is that it is subserved by information channels that do not yield information about anybody's bodily properties except my own (just as introspection does not yield information about anybody's psychological properties except my own). It follows from the simple fact that I somatically proprioceive particular bodily properties and introspect particular psychological properties that those bodily and psychological properties are my own. (1998, p. 149)

I cannot somatically proprioceive that someone's legs are crossed, for example, without thereby proprioceiving that my legs are crossed. And since somatic proprioception is an evidence base for contents where the identification of the subject cannot be in doubt--even for creatures lacking any conceptual resources whatsoever--Bermúdez concludes that nonconceptual states of proprioceptive awareness are IEM.

Focusing just on the particular bodily properties reported on by proprioception, what are we to make of the claim that one cannot be mistaken about which body instantiates certain properties when one is aware of them via proprioception? This question is related to another. To qualify as representational--that is, to be considered contents at all--proprioceptive states must allow for the possibility of misrepresentation. Accordingly, we should consider how can nonconceptual states funded by proprioception misrepresent.

States with conceptual content can, roughly speaking, "who" or "what" misrepresent--they can misrepresent the subject of the state ("who") or the presence of a property ("what"), or presumably both.²⁹ A state that is IEM

²⁹ I limit the discussion to singular thoughts to avoid peripheral complications. Relations can also be misrepresented as can

cannot, obviously, who-misrepresent. Misidentification involves misrepresenting the subject of a state, and immunity as Shoemaker and Bermúdez imagine it can be thought of as a guarantee against self-representational accidents. As we saw above, however, predicative accidents can still occur, which means that such states can nevertheless what-misrepresent. When I express my false belief that I have heard a woodpecker in the distance, I nevertheless succeed in representing myself as the subject of the belief and apparently could not fail to do so. And since states with conceptual content can both misrepresent and be IEM, no questions arise as to their representational status.

Immunity for nonconceptual contents comes at a high price, however. For a state to be representational and IEM it must be possible for it to what-misrepresent without who-misrepresenting. To have what- without who-misrepresentation requires the subject to token a state that successfully represents that subject but erroneously ascribes a property to it. States with conceptual content have an internal structure that permits subject and property representations to succeed and to fail

the objects related. However, given our preoccupation with immunity to error through misidentification, our primary interest lies in the possibilities for subject misrepresentation.

independently. But states with nonconceptual content lack the internal structure requisite for a distinction between what-misrepresentation and who-misrepresentation. For these states to misrepresent at all simply is for them to misidentify. And denying that nonconceptual states can who-misrepresent thus amounts to denying that representational errors of any sort are possible. Hence, endorsing immunity to error through misidentification at this primitive level precludes misrepresentation, which apparently serves to disqualify proprioceptive states--or any nonconceptual states purportedly so immune--from being representational.

Unlike those who discuss IEM judgments, it is not at all clear that proponents of nonaccidental nonconceptual content have the philosophical machinery to relieve this tension. Evans, for example, does not fall into a similar predicament, for his nonaccidental bodily self-ascriptions possess a conceptual structure that localizes--as Shoemaker's condition in its long form indicates--the immunity to error through misidentification relative to the first person pronoun. Misrepresentation can still occur with regard to the predicate position, e.g. where the body is, and hence immunity to misidentification and misrepresentation can co-exist in the same thought or

judgment. Creatures lacking conceptual resources, in contrast, do not have the first-person pronoun or its mental analogue at their disposal. Without conceptually structured thoughts, it seems that these types of subjects cannot possess contents that are both representational and immune to error through misidentification, for they have nothing that that immunity could be relative to.

Or do they? Bermúdez argues that "protobeliefs", or nonconceptual belief analogs, may be ascribed to non-language-using creatures whose behavior requires intentional explanations (1998, pp. 117-119). As he presents them, perceptual protobeliefs³⁰ are nearly as rich as their conceptual correlates. They can embody "nonextensional modes of presentation" in terms of Gibsonian affordances, or the directly perceived possibilities for action and reaction afforded an organism by its environment, such as a location's potential for shelter or nourishment (pp. 121-122). And protobeliefs are somewhat compositional, though they do not allow for the global recombining necessary to satisfy the Generality Constraint (p. 93). So structured, perceptual protobeliefs support primitive inference and the limited generation of

³⁰ Bermúdez also briefly discusses instrumental protobeliefs (p. 118), but our discussion can safely ignore them. Bermúdez draws this bit of his theoretical apparatus from Peacocke (1992).

further new nonconceptual contents from a set of others. Accordingly, perceptual protobeliefs so construed--including contents based on somatic proprioception--seem capable of supporting something like a discrete subject representation that could serve as the locus of IEM as well as a predicative component that could misrepresent a property of the world or body.

One certainly becomes puzzled at this point, however. If nonconceptual contents based upon somatic proprioception can support both a component immune to misidentification and a component vulnerable to misrepresentation, then what are we to make of the original motivation for maintaining a clear conceptual/nonconceptual distinction with regard to contents? Indeed, it seems warranted to think of the constituents of protobeliefs as "protoconcepts". Much like concepts, protoconcepts could be defined in terms of their inferential role, in terms of the protopropositions or protobeliefs in which it participates. As the analogy deepens between concepts and protoconcepts, we have reason to conclude that creatures lacking language do possess conceptual abilities of some sort, however limited or nascent.

Bermúdez himself would no doubt resist this approach since it seems to run afoul of what he calls the the

Priority Principle, or the principle that conceptual abilities require linguistic abilities. Priority was initially important because it "allows us to make a very clear distinction between conceptual and nonconceptual modes of content-bearing representation" (1998, p. 43), and hence provides us with a means of explaining, for example, how conceptual forms of self-consciousness can arise over the course of normal human psychological maturation. Yet, given that protobeliefs are in some measure compositional and fund limited inference--indeed are constituted by protoconcepts--it is no longer clear how we can maintain a sharp distinction between conceptual and nonconceptual contents. And the distinction quickly loses its explanatory force.

Still, perhaps the protoconcept/concept analogy runs fairly shallow, for even if non-language-using creatures possessed a range of protoconcepts defined in terms of protoconceptual roles, they do not have an explicit grasp of these roles. Such creatures are instead merely sensitive to the truth of inferential transitions.

Bermúdez writes:

Certainly, it is possible to be justified (or warranted) in making a certain inferential transition without being able to provide a justification (or warrant) for that inferential transition. It is a

familiar epistemological point, after all, that there is a difference between being justified in holding a belief and justifying that belief. What does not seem to be true is that one can be justified in making an inferential transition even if one is not capable of providing any justifications at all for any inferential transitions. But providing justifications is a paradigmatically linguistic activity. Providing justifications is a matter of identifying and articulating the reasons for a given classification, inference, or judgment. It is because prelinguistic creatures are in principle incapable of providing such justifications that the priority thesis is true. Mere sensitivity to the truth of inferential transitions involving a given concept is not enough for possession of that concept. Rational sensitivity is required, and rational sensitivity comes only with language mastery (p. 71).

For Bermúdez, then, concept possession is a fairly advanced skill based upon an ability to identify and to provide reasons for beliefs, and limited inferential ability--even an ability to make inferences that one is justified in making--does not indicate concept possession.³¹

This seems a bit too stringent, however. Being able to give reasons as reasons is a function of possessing the concepts of justification, belief, and reason, among others. Imposing the further requirement on inferential ability that one recognize that one is in fact giving

³¹ Hurley (1998) and Peacocke (1992) accept similar strong normative and epistemological constraints on concept possession and deployment.

reasons may disqualify attributing conceptual abilities where we normally would be comfortable doing so.³² To take an example Bermúdez himself gives (p. 70), the children in Susan Carey's experiments who concluded that a worm was more likely to have a spleen than a toy mechanical monkey are probably not in position to identify their reasons for this conclusion as reasons and to provide justification for their inferences as justification. Still, he wants to credit these 4-year olds with possessing the concepts HUMAN BEING, LIVING ANIMAL, INTERNAL ORGANS, and the inferential relations between them.

We thus arrive at the conclusion that for a state to be IEM as well as representational it must consist of conceptually structured content. The structural features of a state that bring it into line with the Generality Constraint also underwrite its ability to be about oneself in a way that is IEM. Accordingly, states with nonconceptual content cannot satisfy this condition, regardless of evidence base. For a nonconceptual state to be IEM is for it to be totally immune to representational

³² Circularity no doubt also arises here as it did for Evans. The more a subject has to know about what falls under a concept's extension in order to possess it, the closer one comes to begging questions about possession conditions and what cognitive implications concept possession requires.

error. With no possibility of misrepresentation, however, such states cannot be considered representational at all.

§5. Errors of Bodily Identification

In the previous two sections we considered arguments in favor of the view that certain bodily self-ascriptions are also IEM. The proposals put forward by both Evans and Bermúdez run into largely internal theoretical difficulties. In the next section, I want to consider some empirical work that suggests in a more general way that misidentification remains a possibility when ascribing bodily properties, including limb possession. This work, taken together with the results of the previous two sections, indicates that bodily self-ascriptions cannot be IEM--even when one's senses are functioning normally.

Recall that in discussing Evans we considered a thought experiment in which one could be connected to another's limb via a deviant causal chain in such a way that one's proprioceptive information could not be trusted. However, such cases actually do seem to arise.³³ Oliver Sacks (1985) famously tells of a patient who threw himself out of bed after failing to recognize that the leg he

³³ The recent movie Being John Malkovic cleverly employed this possibility of body switching to great dramatic and comic effect.

encountered under his covers was his own. As Sacks recounts it, the man found, "as he put it, 'someone's leg' in the bed--a severed human leg, a horrible thing!" Concluding that the nurses had played a macabre joke on him, he attempted to throw it out of bed. But, "when he threw it out of bed, he somehow came after it--and now it was attached to him." When told that the mysterious leg was his own, the patient was dumbfounded. Assuring Sacks that he wasn't kidding, he confessed that "a man should know his own body, what's his and what's not--but this leg, this thing ... doesn't feel right, doesn't feel real--and it doesn't look part of me" (pp. 53-54, emphasis in original). One could, it seems, find oneself in the awkward position of needing to identify which limb or even which body was indeed one's own.

Evans could object, of course, that Sacks's patient suffered from a rather extreme form of abnormal sensory function; hence, it provides little insight into the normal case of bodily self-ascription. Perhaps, but misidentification of limbs appears possible even with "normal" subjects. One receives information about one's body from several modalities, including proprioception, touch, and vision. Misidentification of one's limbs can apparently result when separate modalities deliver

conflicting self-specifying information. Botvinick and Cohen (1998), for example, found that normal subjects could be made to experience tactile sensations as within an artificial limb. Subjects were seated at a table with their left arm visually obscured by a screen. A lifelike rubber arm and hand was then placed directly in front of subjects, and they were asked to fix their eyes on the artificial hand. Botvinick and Cohen then stroked the artificial hand and the subjects' actual hidden hand as simultaneously as possible for approximately ten minutes.

Subjects' responses to an open-ended questionnaire about their experience were quite interesting. Subjects reported experiencing the feel of the brush on the artificial hand "as if the rubber hand had sensed the touch" (p. 756). In a follow up test, subjects were subjected to an identical artificial-hand condition as before. This time, both prior to and following stimulation with the brush, subjects were asked to close their eyes and draw their right index finger along an edge beneath the examination table until they believed it was directly below the index finger of their left hand. Botvinick and Cohen found that subjects who had reported experiencing the illusion tended to be off in their judgments of finger alignment rightward towards the artificial hand, with "the

magnitude of this displacement varying significantly in proportion to the reported duration of the illusion" (p. 756).

How are we to explain this strange reaction to the artificial limb? Botvinick and Cohen conclude that "intermodal matching can be sufficient for self-attribution" (p. 756). That is, subjects experienced the tickling sensation as within the rubber arm due to a correlation between proprioceptive and tactile information from the brush stroke on their actual hand and visual information of the activity performed on the artificial hand. In usual circumstances, correlations holding between visual, tactile, and proprioceptive stimulation suffice to locate experiences and to fund accurate self-ascriptions of limb ownership. Unsurprisingly, those subjects in control groups who received brush stimulation on their hidden arm that did not correlate temporally with visual input regarding the artificial arm reported only minor experience of the illusion and did not show reach displacement towards the rubber hand.

Similar results were obtained in a study conducted by van den Bos and Jeannerod (2002) in which the goal was to discover how the visual position of one's hand influenced one's recognition of it. Each of their subjects was seated

opposite an experimenter, and both subject and experimenter extended their right hand in a loose fist. The subject's hand was obscured by an LCD screen that displayed the two hands in one of four rotations. A subject's hand could appear in its normal position at the bottom of the screen (0°); at the right of the screen (-90°); at the left of the screen (90°); or opposite them at the top of the screen (180°). The subject then responded to instructions appearing at the bottom of the screen directing her to extend her index finger, her thumb, or do nothing at all. When the subject was instructed to extend a digit, the experimenter either performed the same movement or its alternative. She was then asked whether the arrow on the screen pointed to the position where her hand had been.

Movement also mattered to successful recognition. Subjects misidentified the experimenter's hand as their own with significant frequency when both made the same movement (either thumb-thumb or finger-finger). And correct attributions of ownership were most difficult for subjects in the no movement condition. This is to be expected--both the subject and the experimenter wore identical gloves, and the subject had no other information upon which to base an identification besides visual position.

Visual position also played a role in limb recognition. Subjects successfully identified their hands most often when they were presented at 0° rotation or how they would normally appear. Trials involving 90° rotation precipitated slightly more errors than -90° trials, but mistakes of attribution were most common in the 180° rotation condition. As we might expect, error rates did not differ significantly in the different movement conditions regardless of rotation. When subject and experimenter made the same movement, however, the 180° rotation condition resulted in the highest number of attribution errors. Interestingly, subjects tended to err on the side of self-attribution, attributing the experimenter's hand to themselves more often than vice versa.

Like Botvinick and Cohen, van den Bos and Jeannerod account for their data in terms of the perceptual cues a subject has available to her to recognize a limb as her own. In Botvinick and Cohen, subjects could appeal to correlations between tactile and visual stimuli to determine limb ownership. In van den Bos and Jeannerod, subjects (depending upon the case) apparently relied upon correlations between proprioceptive and visual stimuli. Just as in the previous study, their data suggest that

subjects base attributions of limb ownership on cross-modal correlations. These correlations are no doubt powerful tools of self-recognition--when subjects' movements differed from the experimenters, they rarely erred in their limb attributions regardless of rotation. Nevertheless, these tools do not guarantee successful limb attribution in all cases.

Studies such as these strongly suggest that even when one's senses--both interoceptive and exteroceptive--are functioning normally, bodily misidentification remains a possibility. We're obviously quite sensitive to correlations between proprioceptive, visual, and tactile stimuli.³⁴ Still, information from these sources does apparently leave room for mistakes of limb ownership, especially when separate modalities support conflicting ownership attributions. Self-ascriptions of bodily properties based on the kind of information singled out by Evans and Bermúdez--including ascriptions of limb ownership--are not, therefore, IEM.

³⁴ And sensitive to it at quite an early age. Work by Bahrick and Watson (1985) suggests that infants as young as five months are sensitive to correlations between their own bodily motion and its visible consequences. See also Bahrick, Moss, and Fadil (1996).

§6. NP and Bodily Self-Ascriptions: The Body Image

The work by Botvinick and Cohen and van den Bos and Jeannerod suggest that normal subjects can indeed misidentify which limb is their own when sensory information from different modalities conflicts or underdetermines ownership. But perhaps the most interesting result was that subjects who misidentified a rubber hand or the hand of an experimenter as their own did not incorrectly guess which limb was theirs. Subjects did not report feeling a sensation in some limb or other and then conclude on the basis of visual information that the rubber limb was the limb in which the stimulation was sensed. Instead, eight out of ten participants in the Botvinick and Cohen study who reported experiencing the stimuli in the artificial hand, for example, also reported "experiencing the rubber hand as belonging to themselves" (p. 756, emphasis added). They offered such comments as "I found myself looking at the dummy hand thinking it was actually my own" (p. 756).

Subjects were compelled to view it as their own despite their awareness that the limb they saw being stroked was in fact artificial. As Descartes himself observes:

Nature also teaches me, by these sensations of pain, hunger, thirst and so on, that I am not merely present in my body as a sailor is present in a ship, but that I am very closely joined and, as it were, intermingled with it, so that I and the body form a unit. (1984, p. 56)

We cannot, therefore, rest content with the conclusion that bodily self-ascriptions are not IEM. We still must explain why they seem nonaccidental to the ascribing subject even when they go awry.

We can find the basis for such an explanation by looking more closely at the conditions which lead to ownership judgments. Interestingly, subjects did not report outlandish ownership experiences. Again, in Botvinick and Cohen, no subjects agreed with the description "It seemed as if I might have more than one left hand or arm" (p. 756). And in replicating the Botvinick and Cohen study, Pavani, Spence, and Driver (2000) obtained similar ownership experiences only when a subject saw the artificial hands aligned with her body as her own arms normally would be. When the hands were visually incompatible with her normal body orientation--when the subject could see where the surrogate limbs ended opposite the fingers, e.g.--her feeling of limb ownership dissipated despite receiving identical stimuli. Farné et al. (2000) report quite similar results. The sense of

ownership a subject felt towards an artificial limb seemed to be "tightly bound to the spatial orientation of the rubber hand" as indicated by vision (p. 2358).³⁵ Instead of experiencing their arms as floating in space or as out of alignment with their body, subjects no longer experienced a sense of ownership towards the artificial limbs at all.

Larger scale ownership effects also seem possible under certain conditions. Cole et al (2000) describe an interesting experience had by those (both normal and impaired) who directed a robot at Johnson Space Center in Houston. Users control the robot's arms via sensors placed along their own limbs that convey their movement and position, and they also wear a helmet that provides video feedback from cameras situated in the robot's "head". Once acclimated to the set up, users report being "in" the robot, so much so that one of the article's authors worried that if he dropped a wrench held in the robot hand it would land on his leg. He felt as if the robot arms were located

³⁵ As with Botvinick and Cohen, subjects of Farné et al. were aware that they were seeing artificial limbs. They write: "... it is important to emphasize that, in our study, the visually compatible and incompatible arrangements of the rubber hand were both completely incongruent with respect to the real position of the patient's right hand. In both cases, this hand was located behind the patient's back" (p. 2358).

directly above his legs where his arms normally would be found.³⁶

These studies suggest that a subject's sense of body and limb ownership depends heavily upon correlations and contingencies between various sensory modalities, primarily touch, proprioception, and vision. But they further suggest that the sense of body ownership relies in large part upon one's beliefs about body organization, including normal limb distribution. Together, they form what Farné et al. call a "peripersonal space" or what other psychologists and philosophers have called a "body image."³⁷ As commonly defined by psychologists, one's body image is "a mental construct or representation, or a set of beliefs about the body" (Gallagher 1995, p. 226). This image includes one's conscious perceptual experience of one's body, one's conceptual understanding of bodies in general (including accepted folk psychological and

³⁶ In "Where Am I?" Dennett (1981) offers a similar thought experiment in which his brain is removed (and backed up by computer) and his body becomes, as it were, the robot. He devises the experiment in which his personhood is sufficiently dispersed to call into question intuitions about the source of one's sense of self and the sense that one is a single unified self. Though we should note that even Dennett himself acknowledges that his scenario tests the bounds of believability.

³⁷ Body image shouldn't be confused with the related by different notion of body schema. The term "body schema" is usually the name given to a set of subpersonal processes that regulate body posture and movement.

scientific conceptions), and one's emotional stance towards one's body.³⁸

We experience bodily sensations such as itches, tickles, and pains as located, but the experienced location is sensitive to our beliefs about how our limbs are normally distributed. Subjects who participated in the Botvinick and Cohen study experienced the tickling sensation of the brush--they experienced themselves being tickled--but under certain conditions they experienced their sensations as immediate and directly in the artificial limb. They did not report first experiencing a disembodied or ambiguously located tickling followed by an attempt to locate it.³⁹ The experience of oneself being tickled and being tickled in a particular location seems intimately tied, and the experienced location is sensitive to our beliefs about how our bodies are organized in terms of their continuity and their spatial distribution.

NP provides an explanation of this remarkable empirical data. Self-ascriptions of bodily properties and of limb ownership are subject to errors of

³⁸ I have the sense that the body image notion as it is used in the literature is a relatively loose one. It's most often invoked in discussions of eating disorders such as anorexia nervosa to explain the difference between how one's body looks to one's owner (fat) and how it actually is (emaciated).

³⁹ Cf. Brewer (1995).

misidentification, but even when one misattributes a limb to oneself one experiences the limb as one's own in virtue of experiencing a sensation immediately and directly in it. No doubt our self-ascriptions of properties and limb ownership based upon tactile stimuli, proprioception, and vision are largely correct. If not, the correlations and contingencies between various sensory modalities could not give rise to a fairly stable body image and a strong sense of body ownership. But since such self-ascriptions rely upon a body image composed largely of a battery of self-specifying beliefs, no doubt the body image can involve conflicts and falsehoods. We believe that our arms project from our shoulders in a certain way, and we discount any limb that looks otherwise.

At first blush results from the van den Bos and Jeannerod study seem at odds with the explanation just given. After all, subjects were able to identify which hand was theirs in all rotations when they were asked to move either their finger or their thumb and that movement differed from the experimenter's. Even when their hands appeared perpendicular to--and detached from--their bodies, subjects had little difficulty indicating which hand was theirs. If attributions of limb ownership were dependent upon a body image in which one's hands and arms are

connected to one's shoulders, why shouldn't errors arise more often in this condition?

This incongruity with NP is only apparent, however. First, given the ubiquity of television, movies, and home video cameras, the subjects who participated in the study were no doubt familiar with video images of themselves and how those images can appear disconnected temporally and spatially from their physical bodies. More importantly, though, errors in ownership attribution increased significantly when subjects and the experimenter made the same movement. Moreover, in the same movement condition, the degree of rotation did matter for error frequency. Mistakes were made most often when the subject's hand was rotated 180°, which means that the experimenter's hand was displayed at 0° or where the subject's hand would normally be. This suggests that when the subject could not uniquely determine which hand was hers by comparing her intention to move and the resulting movement, she relied upon beliefs about where her hand would normally appear visually.

The power of belief to affect ascriptions of limb ownership can be quite striking. Bottini et al. (2002) report the puzzling case of F.B., a 77-year-old female stroke victim. She suffered a right subcortical hemorrhage that resulted in a host of left-sided deficits: hemiplegia

(paralysis of her left side), hemianaesthesia (numbness), and hemianopia (loss of the left half of her visual field). She also exhibited anosognosia, or a denial of her deficits. Unsurprisingly, F.B. did not report receiving any tactile stimulation in her left hand and when asked to touch her left hand she could not locate it, instead randomly exploring the right side of her peripersonal space (p. 249). F.B. displayed only one arresting delusion: She believed that her own left hand belonged to her niece.

This delusion of ownership had remarkable consequences for her reports of touches delivered to that hand. Bottini et al. tested her reaction to left-hand touches in three conditions. She was told prior to the trial that the experimenter would touch (1) her right hand, (2) her left hand, or (3) her niece's hand. F.B. was blindfolded so that she could not see which hand was being touched, and she was instructed to respond "Yes" when she felt the touch or "No" when she did not. Trials were randomized across the three conditions, and Bottini et al. also included "catch trials," ones in which F.B. was told that a touch was forthcoming but no touch was delivered. As expected, F.B. reported touches to her right hand with perfect accuracy, and she denied feeling any touch delivered to her left hand in condition (2).

However, when told that the experiment was going to touch her niece's hand and instead touched her own left hand, F.B. reported feeling the touch with a degree of accuracy much greater than chance.⁴⁰ Following the experiment, Bottini et al. asked F.B. to explain how she could feel touch in someone else's hand. After eluding the question a bit, she eventually replied, "'Yes, I know, it is strange'" (p. 251). Bottini et al. describe her full response as follows:

her absent-minded niece would always forget her hand on the patient's bed while leaving the hospital, so that F.B. used to take care of it until the niece came back for a visit. (p. 251)

Bottini et al. offer a slightly different explanation. They attribute F.B.'s responses in condition (3) to a combination of residual tactile sensory pathways and beliefs about ownership. She could consciously feel touches only when "she delusively referred them to someone else's body image," and in such cases "the disrupted representation of one side of the body is replaced by a preserved image" like that of her niece (p. 252). Accordingly, they conclude that somatosensory stimulation is not sufficient for a sense of limb ownership; it also

⁴⁰ According to the table in Bottini et al. (2002), her accuracy in condition (3) across sets of trials was 70%, 70%, 100%, 80% correct (p. 251).

"depends on a more abstract/higher level body image" (p. 252).

By this point, we have reason to agree with their conclusion. Their steps toward that conclusion, however, need revisiting, particularly regarding the various senses of ownership involved. By justifying her responses in terms of her niece's forgetfulness, F.B. may have been merely rationalizing what she recognized as strange experiences. But in any case if a sense of ownership depends upon possessing a body image, then in referring her conscious left-sided touches "to someone else's body image" as Bottini et al. conclude should preclude her feeling a sense of ownership toward them. Instead, she should report that her niece felt the touch in the niece's left hand. The sense of ownership here is split--F.B. "owns" the experience (experiences herself as being touched) but she also recognizes that her niece owns the hand. Still, we should note that she did not experience the sensation as within, e.g., one of the experimenter's hands. It seems important that F.B.'s niece is forgetful and that F.B. thinks she must "take care of" the hand until her niece returns. Perhaps, then, in taking care of it she treats her niece's normal functioning hand as if it were her own, which could entail noting the experiences it undergoes.

This last peculiar case offers, at best, limited support in favor of NP's explanation of one's sense of body and limb ownership--or perhaps any proposal for that matter.⁴¹ What makes the case especially odd is that her ownership of the experience seems parasitic upon her not owning the hand in which it is experienced in an important sense. And thus far, only F.B. has exhibited this perplexing behavior, and one case is not a sufficiently sturdy base upon which to erect theory. Nevertheless, F.B.'s results, when combined with those discussed above, illustrate the powerful contribution of the body image to one's conscious sense of body ownership. As a result, attributions of ownership can indeed go wrong, which means that such self-ascriptions are not in fact IEM.

§7. Conclusion

We began this chapter with Shoemaker's regress argument in favor of the conclusion that certain self-ascriptions are IEM. For those who find this argument compelling, the regress begins with reflections on the preconditions for

⁴¹ Bottini et al. (2002) really only whet one's appetite for details. One can't help but wonder how F.B. would respond to a host of other questions. When asked to point to her niece's hand, would she point to her own? When her niece was present and she saw her niece's (actual) hand being touched, would she also report feeling a touch? Did these results occur only when she believed her niece's hand was going to be touched or would beliefs about another's hand also suffice?

self-recognition. Once initiated, Shoemaker argues, it must end with self-ascriptions based upon introspection, and hence only certain psychological self-ascriptions are IEM and nonaccidental. Evans and Bermúdez put forward views intending to show that bodily self-ascriptions can also be IEM when based upon certain types of sensory information about one's body. However, we saw that even apart from the theoretical difficulties particular to each view, empirical study of these information channels suggest that errors of misidentification can indeed occur even when they function normally.

We then turned to NP for a more inclusive account of the data. The proper functioning of the senses contributes to an overall body image consisting of a set of beliefs and attitudes that together constitute a sense of ownership. NP nicely accommodates the notion of body image as a subset of one's battery of self-specifying beliefs that serve to identify for the subject which body belongs to her and to ground her sense of body ownership. The body image, combined with cross-modal correlations in sensory input, gives rise to a powerful sense of limb ownership. To feel a sensation in a limb is for one to seem to feel it in one's own limb, but mistakes about whether the limb in

question is one's own can nevertheless arise. Bodily self-ascriptions at best seem nonaccidental.

Where does this leave us? Evans fears that if bodily self-ascriptions fail to be IEM, our nonaccidental thoughts about ourselves would be hospitable to Cartesianism. The fear does not seem misplaced. If introspective self-ascriptions are the only ones that can stop the regress of identification, then we do seem to have a fertile ground of difference between mind and body in which Cartesianism can take root. Moreover, as we saw at the outset, Shoemaker contends that self-ascriptions of bodily properties involve identifying, for example, which body responds to my intentions to move. This tempts one to think that one can't be wrong about whether oneself is intending such-and-such, but one can be wrong about which body is one's own.⁴²

Disagreeing with Evans and Bermúdez does not mean agreeing with Shoemaker and endorsing Cartesianism, however. Accordingly, in the next chapter we turn to psychological self-ascriptions based on introspection. As we go forward, we should keep in mind the lessons learned

⁴² It's worth noting that van den Bos and Jeannerod describe their results as demonstrating that one often recognizes oneself in virtue of "the matching of one's intentions and the bodily effects of self-generated actions" (p. 178). That is, they assume that subjects are aware of their own intentions and that errors result from mismatching those intentions with the bodily effects of others or other things.

in accounting for the errors one can commit when self-ascribing bodily properties and limbs and the accompanying sense of body and limb ownership. One's sense of ownership may seem immediate and direct, but this appearance does not mean that self-misidentification is impossible. Bermúdez was right insofar as the bodily properties reported on by proprioception seem to be one's own, but even in the normal case they can also seem to be instantiated within a limb that does not belong to oneself. How things seem does not settle the matter as to how things are, but any satisfactory account of self-ascription must speak to both. We saw that NP was able to explain how things are with regard to bodily self-ascriptions without misplacing their seeming to be IEM. The challenge ahead is to do demonstrate that it can do the same in answering the Misconception Problem, the Amnesia Objection, and the Asymmetry Problem introduced in Chapter II. We now turn to that final task.

Chapter V. Misidentifying the First Person

"My intention was always to get away from myself, though I knew perfectly well that I was using myself. Call it a little game between 'I' and 'me'."

--Marcel Duchamp, explaining
readymades¹

§1. Introduction

We arrived at the end of Chapter IV with a success and a difficulty. The success was an explanatory one: NP provided a means of accounting for both the appearance of IEM with regard to a certain class of bodily self-ascriptions and the actual errors one is subject to even when one's senses function normally. Proposals that attempt to extend the class of IEM self-ascriptions to those involving bodily states and properties, like those offered by Evans and Bermúdez, at best can do only the former. Such proposals take how things seem to us to settle the question as to the way things are, but doing so precludes an explanation of a range of compelling empirical data regarding the fragility of one's sense of bodily ownership.

The difficulty, however, falls out of this explanatory success. As we saw in previous chapters, NP

¹ Quoted in Tomkins (1996, pp. 159-160).

excels at accounting for errors of self-misidentification, and since bodily self-ascriptions are in fact subject to such errors, one might protest that no significant progress has been made. Indeed, the discussion of bodily ownership seems to push one towards a Cartesian conclusion--namely that mistakes of ownership can occur because owners are in some important sense separate and therefore separable from what they own.

Accordingly, Shoemaker's neo-Cartesian explanation of these mistakes acquires a much stronger pull. For him even those errors of misidentification possible with bodily self-ascriptions require that one already possess identifying beliefs about one's psychological states and their behavioral implications, beliefs that are themselves not based upon self-identification of any sort. Moreover, one might be tempted to think, as Shoemaker does, that one can be aware of one's psychological states in a way that gives rise to IEM self-ascriptions, whereas one has no such awareness of one's bodily states and properties.

We can frame the puzzle in a slightly different way that more directly confronts NP. We found in Chapter IV that to ascribe a property F to a familiar individual a by identifying some presented b as a seems to require that one already possess beliefs about the properties that a

instantiates, properties that overlap at least in part with the properties one believes b to possess. Regress threatens, however, if we further hold that one ascribes to a what one takes to be its identifying properties on the basis of further beliefs about some individual c. One's battery of self-specifying belief may open up the possibility of misidentification because of its identificatory role, but those beliefs themselves apparently cannot be based upon self-identification. Hence, in explaining the features of bodily self-ascriptions, NP hasn't fully addressed the relationship between one's battery of self-specifying beliefs and the self-ascriptions that draw upon it. As a result, its success seems hollow at best.

Though for these reasons we may be quite tempted to endorse Shoemaker's view that self-ascriptions of introspected psychological states and properties must be IEM, I believe it to be mistaken. Just as one appeared to be directly aware of certain bodily properties as one's own, we will find that an appearance/reality distinction exists with regard to psychological self-ascriptions based upon introspective awareness. For empirical work strongly suggests that, appearances to the contrary, we have no such direct and unmediated access to our mental lives. To

motivate that claim, in §2 I consider the importance of consciousness to certain types of self-ascription--not often considered in discussions of IEM--and how our mental lives seem to us even apart from how they actually are. From the point of view of consciousness it does seem to us that one can't be mistaken whether oneself is in a particular mental state when one introspects. But this appearance results not from some special introspective access to our mental lives but is instead an artifact of consciousness, a result of conscious interpretation. Though proponents of IEM could object that a vestige of immunity remains in that one still can't be mistaken about who introspects a particular state or property, we see in §3 that self-interpretation extends to interpreting the introspecting and ascribing subject as well. Under NP the identified self is, as it were, constituted by that battery. The self of accidental and nonaccidental self-ascriptions alike can be thought of, in Dennett's parlance, as a center of narrative gravity. Ultimately, the intuitions regarding the apparent need for IEM self-ascriptions can be traced back to how our conscious lives appear to us.

Giving proper place to consciousness and the role of self-interpretation in introspective self-ascriptions, in

§4 I review the development of the dialectic demonstrating that NP provides a satisfactory, comprehensive account of self-ascription in light of our original concerns. With the full proposal in hand, I turn in §5 to the Amnesia objection and to the Misconception Problem. Finally, in §6 I explain away the Asymmetry Problem in terms of the richness of one's nonconscious battery of self-specifying beliefs. Nonaccidental self-ascriptions, canonically expressed with the first-person pronoun, do not therefore differ in kind from accidental ones. As a result, self-misidentification always remains a possibility, even when self-ascribing psychological states and properties. I conclude in §7 with a few outstanding questions and some avenues of further research.

§2. Consciousness, Introspection, and Self-Interpretation

As we saw in Chapter III, Shoemaker argues that what makes a particular self-ascriptions IEM is the way in which one is aware of the property or state ascribed (1968, p. 565).² One does, after all, self-ascribe properties and states in a variety of ways, and on occasion we even become aware of our own mental states fairly indirectly. A close friend or

² And according to Evans and Bermúdez, the importance lies in the evidence base or type of information funding a particular self-ascription.

spouse can bring one's anger to one's attention before one realizes it oneself, for example. But this type of psychological self-awareness doesn't rule out errors of self-identification. IEM is supposed to capture something unique about certain self-ascriptions that distinguishes them in kind from those made on the basis of third-person information. Accordingly, for Shoemaker, a self-ascription is IEM in virtue of one's being introspectively aware of the mental states and properties ascribed.

As we saw in Chapter III, Shoemaker argues that mental states are self-intimating, which is to say that "it is of the essence of many kinds of mental states and phenomena to reveal themselves to introspection" (1996, p. 242). Recall also that Shoemaker explains introspective access in functional and conceptual terms. If a creature believes that *p*--and assuming that it has a similarly expressive conceptual repertoire and a level of rationality rivaling ours--it will behave as if it were aware of its believing that *p* in the same way that we are introspectively aware of our own states. One's introspective self-ascriptions are therefore caused by the states they ascribe to oneself, and this causal relation is constitutive of first-person introspective access. And because such introspective self-ascriptions involve no

self-identification component, self-misidentification cannot occur.³

To this point we have said little about how state consciousness figures in IEM self-ascriptions. Although he rarely discusses IEM in these terms, it's clear that Shoemaker believes that a particular state or property must be conscious for its self-ascription to be IEM. To take one representative remark, Shoemaker contends that "if I have my usual access to my hunger," then the question "Someone is hungry but is it I who is hungry" makes no sense (1996, p. 211, emphasis added). By "usual access" Shoemaker means one's apparent direct and unmediated access to one's hunger--or as we might commonly say, how we usually feel when we report our hunger. When one introspects a mental states or property, it becomes conscious, where, according to Shoemaker, "a mental entity's being conscious involves its revealing its existence and nature to its possessor in an immediate way" (1996, p. 224). The claim is that when my hunger is

³ As we noted above, Shoemaker believes that one's mental states and one's beliefs about them are not, in fact, distinct. Instead, their relation is one of core realizer to total realizer. But as we also noted above (and will note again shortly), introspective ascriptions can be distinct from the states ascribed in that one can be in the state without being directly and immediately aware of it and one can be introspectively conscious of oneself as in a state that one is in fact not in.

conscious I can't be mistaken about whether it is I who is hungry, whoever I may be.

For a mental state to be conscious one must be aware of it in a direct and unmediated way, and according to Shoemaker all IEM self-ascriptions result from this type of self-awareness. Therefore, all IEM self-ascriptions must be conscious. Still, we must mark the difference between states that are conscious and those that are introspectively conscious. Rosenthal (2000) draws this distinction in the following way:

When we introspect a state, we are conscious of it in a way that seems attentive, focused, deliberate, and reflective. When a state is conscious but not introspectively conscious, by contrast, we are conscious of it in a way that is relatively fleeting, diffuse, casual, and inattentive. Introspective and nonintrospective consciousness do not seem to differ in any other ways. There is no other phenomenological difference, and no theoretical reason to posit any additional difference. (p. 208)

The two differ, then, in the disparate ways in which one can be conscious of the same state.

No doubt this is an important respect in which the two differ, but we should note another. Intuitively, when one is introspectively conscious of one's mental states, one seems not only aware of the state introspected in a direct and unmediated way, but one also seems to be

similarly aware of oneself as its bearer. And this difference impacts the speech acts one performs about oneself. We often express our conscious states with speech acts that do not involve the first-person pronoun or any self-designating expression, whereas introspective self-ascriptions report our conscious states. These reports--usually of the form "I am F" or "I believe that p," for some sensory state F or for some intentional state p⁴--are themselves expressions of conscious states. The property or state is conscious in both cases, but in introspective consciousness one is also conscious of one's being in that conscious state.⁵

As Rosenthal (2003) notes, this intuitive awareness we have of ourselves in introspection seems rather thin. One seems to be aware of oneself as the bearer of the introspected conscious mental state, present only in the

⁴ Here again I intend the remark to cover all propositional attitudes.

⁵ Rosenthal (2003) does note something like this difference. "Introspective consciousness occurs," he writes, "when we are not only conscious of [our mental] states, but also conscious that we are" (p. 332).

Hume's famous remark regarding the elusiveness of the introspected self may spring to mind at this point as well. After all, he claims that when he "looks inward" he observes at most a parade of perceptions but no perceiver (1978, p. 252). But as we saw in Chapter III, Hume's mistake lies in taking the word 'introspection' literally and adopting a perceptual model of introspective self-awareness. When we introspect, we don't cast an "inner eye" upon a mental landscape and ascribe properties to whatever subject we "see" lurking there. Nevertheless, introspective self-ascriptions do express our intuitive awareness of not only a particular mental state but of the bearer of that state.

act of thinking. And this apparent direct and sparse awareness of oneself no doubt gives rise to the impression that the introspecting subject cannot misidentify the individual whose states and properties are introspected. While Bermúdez's claim may have been misleading with regard to proprioception, it certainly does seem that "introspection does not yield information about anybody's psychological properties except my own" (1998, p. 149). From the point of view of consciousness, the state introspected and the introspecting subject are inseparable.

Though weak self-intimation would indeed explain why certain self-ascriptions are IEM, we found in Chapter III that it does not survive scrutiny. A self-blind creature--one that completely lacks characteristically first-person or introspective access to her own states--cannot be ruled out on conceptual grounds. Indeed, most of our intentional states are probably not conscious, and the fascinating phenomenon of blindsight also suggests that one can be in mental states with qualitative character without those states being conscious.⁶ Hence, we can be in states without being introspectively conscious of them.

⁶ See Chapter III, §5. Studies involving masked priming also suggest that sensory states or states with qualitative character can be nonconscious. In these studies subjects are shown stimuli for only a few milliseconds. They report seeing nothing, which indicates that they did not consciously see the stimuli. Nevertheless, subsequent

More important for present purposes, however, empirical work in psychology strongly suggests that one can mistakenly believe that one is introspectively conscious of a particular mental state that one could not be in. That is, one can self-ascribe mental states and properties on the basis of one's apparent direct and unmediated awareness of them yet not be in those states. For example, one can apparently be mistaken as to one's conscious qualitative states. Rosenthal (2000) cites the example of apprehensive dental patients. During dental work on a tooth that has been anesthetized or lacks a nerve altogether, certain patients will nevertheless report feeling pain in that tooth. One possible explanation, and the one that Rosenthal favors, is that the fear such patients have about the procedure and its expected painfulness causes them to misconstrue vibrations resulting from the drill as sensations of pain. Once the patient has been given this explanation (and reminded of her tooth's numbness), it seems subjectively to her that she is fearful and that the drill is causing her tooth to vibrate, but she no longer experiences pain. Nevertheless, it still seems to her that

behavior such as word choice and completion is affected by what was flashed before them. For representative work of this sort, see Marcel (1983a) and (1983b).

she was originally in pain, despite later realizing its unlikely source.

We make similar and arguably more frequent mistakes about our intentional states. Perhaps the most famous source of support for this claim (one briefly mentioned in Chapter III) is due to Nisbett and Wilson (1977). In one of the many studies they discuss, subjects were told that they were participating in a consumer survey. They were asked to examine articles of clothing--four different nightgowns in one trial and four identical pairs of nylon stockings in another--and rank them according to quality. Subjects were then asked for their reasons behind their rankings. Interestingly, though all four pairs of stockings were in fact identical, subjects overwhelming tended to favor the right-most pair. No subject justified her choice in terms of article position, however, and when subjects were asked directly whether position had an effect on their choice, "virtually all subjects denied it, usually with a worried glance at the interviewer suggesting that they felt either that they had misunderstood the question or were dealing with a madman" (p. 244).

Nisbett and Wilson explain their results in terms unfavorable to introspection as a special form of self-

awareness.⁷ As the title of their paper indicates, subjects often seem to "tell more than they can know" about their own mental states and processes. Instead of introspecting what states and processes actually led to their choices, Nisbett and Wilson suggest that when asked, subjects:

did something that may have felt like introspection but which in fact may have been only a simple judgment of the extent to which input was a representative or plausible case of output. (p. 249)

Instead of accurate self-reporting, what we call introspection is actually a process in which we often confabulate reasons for what we do based upon widely held theories about the causal relations between intention and action. Those confabulated reasons can seem, from the point of view of consciousness, to be the ones we are directly and immediately aware of having.

Subjects can also mistakenly have the sense that they consciously willed a particular action when in fact they did not. Wegner (2002) argues that one's experience of oneself as a causal agent arises not from introspective access to one's intentions to act but rather from

⁷ The clothing trial was but one of many they conducted, and they also discuss a fair amount of similar work in this seminal paper. This type of research has also continued apace. For a more recent survey and extension of this work, see Wilson, Hodges, and LaFleur (1995).

inferences regarding the priority, consistency, and exclusivity of the thought responsible for the action (p. 69). Though it may seem that a conscious thought about an action causes it, Wegner suggests that the thought and the action itself may have a common, nonconscious cause. Accordingly, in simply taking into account one's conscious sense of will "it would be impossible to tell in any given case whether your thought was causing your action, or something else was causing both of them" (2003, p. 65).

Under this picture, it should be possible to fool subjects into experiencing themselves as consciously willing an action by manipulating their beliefs about their own thoughts. This indeed seems possible. Wegner and Wheatley (1999) conducted what they called the "I Spy study" (p. 74), which focused upon the role of an individual's belief about the priority of her intention to act in her experience of will. Each participant in the study arrived at the lab at the same time as a "confederate" posing as another test subject. The two were seated facing each other, with both hands placed on top of a modified computer mouse (much like the planchette of a Ouija board, the inspiration for the paradigm). Together, subject and confederate were asked to move the pointer

around a computer screen that showed a photo of about fifty small objects.

Subjects were told that the study's aim was to investigate individuals' feelings of intention. The subject and confederate were told to stop moving the mouse every thirty seconds and then rate each stop according to the degree in which it resulted from their own intentions to stop it. The scale ranged from 0% ("I allowed the stop to happen") to 100% ("I intended to make the stop"). Participants were also told by the experimenter that they would wear headphones during the experiment and that each 30-second interval would be punctuated by 10 seconds of music indicating that they should stop moving the mouse. Both subject and confederate were told that they would be fed different soundtracks including a different word each trial and that any words they heard were only intended to distract them. (To complete the ruse, the experimenter even played a bit of the tape for each and asked them what word they heard--the confederate always reported hearing a different word than the actual subject.) Once the trials began, however, subjects heard a word priming them to choose an object in the picture while the confederate heard specific instructions about what movements to make with the mouse and when.

Interestingly, the timing of the confederate's movements greatly affected the subject's rating of the stop. In most of the trials, the confederate heard no word and simply let the subject stop the mouse. In four "forced-stop" trials, however, the confederate heard instructions to move the pointer to an object on the screen corresponding to the word heard by the subject, along with a countdown to stopping the movement. Wegner and Wheatley varied the time between the heard word and the confederate's causing the pointer to stop on the corresponding picture so that the subject heard the word either 30 seconds before, 5 seconds before, 1 second before, or 1 second after the stop. Subjects rated their intention to stop the mouse as most effective in the 1-second trials and only slightly less so in the 5-second trials. In contrast, subjects experienced little personal intention in both the 30-second and 1-second-after trials. In fact, the mean percentage reported on the unforced stops (where subjects had total control of when and where to stop) was roughly the same as that for the 1- and 5-second forced-stop trials (Wegner 2000, p. 78).

Much like Nisbett and Wilson, Wegner (and Wegner and Wheatley) explains the results as an artifact of beliefs about mental causation. We believe that intentions can

only cause actions if they immediately precede them. The notable thing is that subjects experienced themselves as intending to stop the mouse even when they were not the cause of it and so did not have the requisite intentional state to do so. "Apparently, the experience of will can be created by the manipulation of thought and action in accord with the principle of priority," writes Wegner, "and this experience can occur even when the person's thought cannot have created the action" (2000, p. 78). Beliefs about our states and their relation to actions can themselves give rise to the subjective experience of willing something to occur even when we are not in those states.⁸

Perhaps the most extreme and puzzling illustration of introspective error is the delusion of thought insertion.

C. D. Frith (1992) offers a striking example of the phenomenon as recounted by a schizophrenic patient:

Thoughts are put into my mind like "Kill God". It's just like my mind working, but it isn't. They come from this chap, Chris. They're his thoughts. (p. 66)

Mellor (1970) provides an additional example but with inserted emotions:

The patient experiences feelings which do not seem to be his own. The feelings are attributed to some external source and are

⁸ For an overview of this study and others relevant to consistency and exclusivity, see Wegner (2003).

imposed upon him. A 23-year-old female patient reported: "I cry, tears roll down my cheeks and I look unhappy, but inside I have a cold anger because they are using me in this way, and it is not me who is unhappy, but they are projecting unhappiness into my brain. ... You have no idea how terrible it is to laugh and to look happy and to know that it is not you. (p. 17)

Certain subjects suffering from schizophrenia report episodes in which they experience a particular thought or feeling as occurring within their own stream of consciousness but view it, quite literally, as authored by another.

It's difficult to take such subjects at their word regarding how their mental lives actually are--no doubt the "inserted" episodes do not in fact have external tokeners. But it's equally difficult to explain their sincere (dis)avowals in terms of direct and unmediated access to their own states. If they did have such access, these alien states should subjectively seem like anything but alien, not only states that they possess but states that they authored as well.

Though much work remains to provide a satisfactory and comprehensive explanation of thought insertion, Stephens and Graham (2000) offer the promising beginnings of one not based upon introspection as traditionally conceived. Without delving too deeply into the

peculiarities of their view, they contend that a subject experiencing thought insertion "will not accept as agentically her own thoughts whose occurrence she finds inexplicable by reference to her conception or self-referential description of her intentional states" (2000, p. 170). That is, a subject denies that she thinks a particular thought that she is directly and immediately aware of because she does not believe that she is the kind of person who would think such a thought. Using Frith's example (and simplifying a bit),⁹ the troubled subject who finds the thought "Kill God" "popping" into her mind, doesn't see herself as one who would likely commit deicide. Since the thought nevertheless seems intelligible, she seizes upon an alternative explanation in which an "external" agent has authored it and has forced it into her mind.¹⁰

⁹ Frith's own explanation of schizophrenic delusions is similar to that of Stephens and Graham in that he attempts to provide a fundamental account of both auditory hallucinations and delusions of thought in which agency plays a central role. For him, these afflictions result from a breakdown in the metarepresentational systems that monitor our intentions to act. Stephens and Graham find Frith's account incomplete in that it fails to account for why a subject takes a particular thought to be the result of some external force and therefore cannot maintain the clinical distinction between thought influence and thought insertion (2000, pp. 141-42).

¹⁰ Stephens and Graham expressly offer that they do not want to be bound to any particular account of why a subject might view a thought as repugnant given her psychological self-conception, for "perhaps on some occasions a person fails to find an intentional explanation for a given thought because there isn't one" (2000, p. 170).

Milder examples of this type of introspective anomaly can be found outside the clinic as well. Recovering from injuries sustained while hiking, Oliver Sacks catches himself reacting with unexpected intensity to watching a team practice rugby:

I was surprised and appalled at the spasm of hate in myself. ... I looked at them with the virulent envy, the mean rancor, the poisonous spite of the invalid; and then I turned away; I could bear them no longer. Nor could I bear my own feelings, the revealed ugliness in me. I consoled myself by saying, "This isn't me--the real me--but my sickness which is speaking." (1984, pp. 176-7)

Sacks's vivid account has much in common with those suffering from thought insertion.¹¹ He does admit that the "poisonous spite" that visited him revealed something about himself. But the thoughts and experiences that Sacks is aware of in a direct and unmediated way seem so out of character that he feels forced to find consolation in denying ownership of them. It proves easier to think of them as not his own reaction to his current state but as the product of a surrogate "speaker"--his injured state itself.

¹¹ In fact, Maher has consistently argued (1974, 1988a, 1988b, and 1999, e.g.) that reports by subjects suffering from delusions such as thought insertion are trying to make sense of experiences like Sacks's only more extreme. That is, the source of the delusion lies not in irrational interpretation but sincere interpretation of extremely anomalous experiences that differ only in degree from our own.

How should we make sense of this distinction at the level of introspective consciousness between how our mental lives can appear to us and how they really are? In light of empirical work and more common experiences like that discussed above, Dennett suggests an alternative to the traditional view of introspection as special self-access:

We postulate all these apparent activities and mental processes [as captured in folk-psychology] in order to make sense of the behavior we observe--in order, in fact, to make as much sense as possible of the behavior, especially when the behavior we observe is our own. Philosophers of mind used to go out of their way to insist that one's access to one's own case in such matters is quite unlike one's access to others', but as we learn more about various forms of psychopathology and even the foibles of apparently normal people (see Nisbett and Wilson 1977), it becomes more plausible to suppose that although there are still some small corners of unchallenged privilege, some matters about which our authority is invincible, each of us is in most regards a sort of inveterate auto-psychologist, effortless inventing intentional interpretations of our own actions in an inseparable mix of confabulation, retrospective self-justification, and (on occasion, no doubt) good theorizing. The striking cases of confabulation by subjects under hypnosis or suffering from various well-documented brain disorders (Korsakoff's syndrome, split brains, various "agnosias") raise the prospect that such virtuoso displays of utterly unsupported self-interpretation are not manifestations of a skill suddenly learned in response to trauma, but of a normal way unmasked... (Dennett, 1987, p. 91)

Instead of viewing introspection as a kind of direct access to the mental states and properties that we actually have, we should consider it a form of self-interpretation masquerading as a special form of privilege.

Rosenthal agrees. Like Dennett, he argues that we have reason to believe that introspection amounts to self-interpretation, even when our introspective self-ascriptions of mental states prove correct. However, he stresses that subjects nevertheless take themselves to have introspected the confabulated states, and it seems to them from the point of view of consciousness that they are directly and immediately aware of themselves as being in those states:

People interpret themselves in the light of their situation and past experience, and some of these self-interpretations have to do with what mental states they are in. As long as one remains unaware of whatever inference and motivation leads to these self-interpretations about one's mental states, the self-interpretations will seem, from a first-person point of view, to be spontaneous and unmediated. They will seem to arise from just asking oneself what mental states one is in, from a deliberate decision to focus on the states in question by casting one's mental eye inward. But it is likely that such introspective awareness results in substantial measure from desires to see ourselves in a certain light. Introspection is often, if not always, a process of conscious self-interpretation. (2002, p. 224)

However directly and immediately our awareness of own states might subjectively seem, experimental and folk psychology teach us that we have no such awareness.

In the absence of such special awareness, self-interpretation amounts to explaining or predicting our own behavior in much the same way as we explain and predict the behavior of others. To provide satisfactory explanations and to make successful predictions we adopt an interpretive strategy that Dennett (1987, 1991) calls the "intentional stance." We have a wide range of strategies at our disposal to predict and to explain the behavior of a particular individual or system. We can explain chemical reactions or the interactions of bodies in space, for example, by taking the physical stance toward them--that is, by subsuming those particulars under physical theories that deliver predictions and postdictions in light of initial conditions. Despite its descriptive power, not all individuals or systems can be easily explained from this stance. For certain systems, we choose to adopt what Dennett calls the "design stance," or the strategy where one predicts future behavior by considering how an individual was designed to behave (1987, pp. 16-17).

Whereas the physical stance trades in subatomic particles, forces, and the laws that govern them, the

design stance operates on a more familiar level of abstraction. We can easily manipulate fairly complex machines like telephones and automobiles without knowing much if anything about how they actually work, let alone their molecular makeup. We can effortlessly predict the outcome of pressing a certain sequence of numbers on a telephone by relying upon assumptions about how the device is designed to fulfill its communicative function--e.g., that it is connected in a suitable way to other similar devices, that a reliable relation exists between a sequence of numbers and the ringing phone in one's mother's house, that one speaks into one end of the handset and listens to the other. Those who repair phones no doubt understand them better, but they do so by having a more thorough grasp of their design.

We can greatly increase our predictive power regarding intentional systems like ourselves by adopting the intentional stance, a predictive strategy based upon the ascription of beliefs and desires to the individual in question. Dennett instructs us in adopting the intentional stance as follows:

first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out

what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many--but not all--instances yield a decision about what the agent ought to do; that is what you predict the agent will do. (1987, p. 17)

Dennett further argues that we employ this strategy so often and with such success that we take it for granted. Our expectations of our coevals are so often met that the occasional inappropriate utterance or unpredictable behavior stands out in stark relief.¹²

We also adopt the intentional stance towards ourselves, ascribing to ourselves various beliefs, desires and their interactions. As a result our own behavior can flout expectations we may have of ourselves. No doubt Sacks would normally deny that he is capable of such a virulent reaction to such a harmless scene, and his

¹² The lion's share of the discussion surrounding Dennett's proposal concerns whether endorsing the intentional stance commits one to instrumentalism about mental states. If we (or some other creature) could come to predict our behavior as effectively from the physical stance, one argument goes, does that mean that we, like simple chemical compounds, don't really have beliefs and desires? (See, e.g., Dennett, 1987, pp. 25-6.) Dennett's handling of this objection hasn't helped matters much, for he wants to "place [his] view on the knife-edge between the intolerable extremes of simple realism and simple relativism" (1987, p. 38). This edge affords little room to balance, for he argues that the patterns of behavior are real, but they can only be discerned and explained at a particular level of abstraction. We can remain safely neutral on this issue, however, for one could construe the success of that predictive strategy as Stephens and Graham do--namely, as solid theorizing about our actual "underlying, relatively more persistent intentional states" (2000, p. 161).

surprise results from observing a certain "ugliness" befitting a sickness and not his "real" self.

Anomalies prove fairly rare, however, and our own experiences and behavior usually confirm our predictions. As a result, we find ourselves drawn to the picture of introspection as something like direct inner perception. Dennett nicely unravels this intuition by comparing it to Hume's diagnosis of our intuitions about the necessary connection between causes and their effects. Though it appears to us that we expect particular effects given particular causes because we have observed a necessary connection between causes and their effects, Hume observed that this picture gets things backwards. Since effects are logically distinct from any cause, our expectations of necessary connection must instead be borne of habit, the result of observing constant conjunction between two events. As a result, we find ourselves drawing inferences from cause to effect and so are taken in by the illusion that a necessary connection grounds that inference. Similarly, Dennett argues that "we find ourselves wanting to say all these things about what is going on in us; this gives rise to theories we hold about how we come to be able to do this--for instance, the notorious but homespun theory

that we 'perceive' these goings on with our 'inner eye' ... " (1978, p. 166-7).

The above observations concerning introspection and our intuitions about it square nicely with NP. For according to it, when we ascribe a mental state or property we draw upon a rich battery of self-specifying beliefs. This battery provides the basis for explanations of our past and expectations of our future behavior. Introspection thus amounts to a type of informed, conscious self-interpretation. The battery constitutes a kind of predictive theory about ourselves that takes as its evidence base one's collective behavior, both verbal and non-verbal. Accordingly, introspective self-ascriptions do not express a special kind of access to one's mental life but rather amount to explanations of one's behavior in light of the battery of self-identifying beliefs one possesses. When we find ourselves behaving in a way that runs counter to the ways in which we identify ourselves, we face falsified predictions about our own behavior, and more often than not (though not always as Sacks's example suggests) we revise the battery to explain why we did or said what we did.

Only an account that no longer places stock in introspective privilege can accommodate the confabulations

and surprises not uncommon to our mental lives that empirical research reveals. These considerations show that, appearances to the contrary, we have no special self-access of the requisite sort that those like Shoemaker appeal to to explain IEM. Consequently, a comprehensive theory of self-ascription has no room for this notion.

§3. Who's Interpreting Whom: Self-Interpretation, Self-Identification, and Unity of Self

We have seen that introspective self-ascriptions do involve self-identification in that they amount to explanations of behavior in light of the battery of self-identifying beliefs. Still, proponents of IEM would no doubt object that we have not quite earned the dramatic conclusion of the previous section. Shoemaker introduces IEM as a technical notion to make sense of the theoretical need for self-ascriptions that must not, on pain of regress, involve self-identification. By explaining the relation between the battery of self-specifying beliefs and self-ascription as one between theory and the predictions and explanations it yields, we remain open to the charge that we still have not completely disarmed the regress of identification.

The charge can be sharpened in the following way. It can't, as it were, be interpretation all the way down. Characterizing introspection in terms of identification and

interpretation seems to presuppose an individual that is both the subject and object of the interpretation. After all, doesn't interpretation require someone to interpret and someone to be interpreted? One may identify oneself via a battery of self-specifying beliefs, but can't one still argue that a vestige of IEM exists in that a question can't arise as to whether oneself is both interpreter and individual interpreted? Can one interpret another as oneself? If so, doesn't the regress motivating IEM self-ascriptions still loom menacingly? Who's adopting the stance toward herself? Who's interpreting whom?

It's difficult not to be gripped by these questions even at this point in the overall discussion, and this illustrates the persistence and power of our intuitions about self-ascribing states and properties. Nevertheless, intuition does not always prove a reliable guide to facts of the matter. The empirical evidence adduced in the previous section also suggests that the self-interpretation constitutive of introspective consciousness not only leads to self-ascriptions regarding what mental states we have but also whether oneself is the subject who has them. The self is a product of the battery--a fictional posit useful for making larger sense of our behaviors and experiences--and not the precondition for possessing it. Once again,

the appearance of a substantive self to which properties are ascribed is an artifact of consciousness.

Consider the claim motivating IEM. Certain self-ascriptions are not subject to errors of misidentification because one can be aware of the property or state ascribed in a way that ineluctably makes one aware of the possessor of that property or state. For me to be introspectively aware that *p* is believed is to be aware that I believe that *p*. But, appearances notwithstanding, if we actually have no such direct and immediate access to our psychological states, why should we think that we have direct and immediate awareness of the possessor of those states? If we confabulate the ascription, why can't we confabulate the ascriber?

The suggestion certainly sounds a bit like a joke. What could it mean to confabulate the ascriber--for one to confabulate oneself? If anything at all is obvious to me, it's that I am no fiction.

In one sense, selves certainly are real. We, like all creatures, make a functional distinction between self and world. Our biological boundaries are lines of defense, keeping out many would-be invaders and sequestering and protecting the invaluable elements that sustain us. These biological boundaries are not hard and fast, however.

Ruminants, much like us, rely upon bacteria to help break down ingested plants for digestion, spiders build webs, birds build nests, hermit crabs seek out and occupy discarded shells. And at times, our own systems can turn against our own normally welcome organs and cells to form tumors.¹³

Dennett (1991, 1992) argues that as language users we have our own unique way of building upon our biological boundaries. Instead of spinning webs or appropriating shells, we appropriate words and weave them "like spiderwebs into self-protective strings of narrative" (1991, p. 417). And we do it in such a way that promotes self creation:

These strings or streams of narrative issue forth as if from a single source--not just in the obvious physical sense of flowing from just one mouth, or one pencil or pen, but in a more subtle sense: their effect on any audience is to encourage them to (try to) posit a unified agent whose words they are, about whom they are: in short, to posit a center of narrative gravity. (1991, p. 418)

Just as billions of independent cells coalesce to form the biological self, countless narrative "cells" clump together

¹³ As Dennett (1990, p. 414) notes, these biological boundaries have psychological repercussions as well. In fact, to quote Miller (1997), "[d]isgust rules mark the boundaries of self; the relaxing of them marks privilege, intimacy, duty, and caring" (p. xi). Indeed, he devotes his fascinating and extensive work to mapping the topography of disgust largely in terms of our reactions to biological border crossings.

to constitute narrative selves. The resulting centers of narrative gravity are thus useful fictions for organizing a complex pattern of behavior, including the self-ascriptions and self-interpretations that we are prone to make.

Dennett argues that we're not only masterful pattern generators, but we're accomplished pattern trackers as well. And it's good that we are, since we put these fictional posits to great use in keeping track of and preserving our minimal biological selves. Hungry creatures must avoid eating themselves, and they must also avoid enemies and pursue sustenance and mates. This requires that one be sensitive to which behavior is one's own--and which is not.

Our means of determining which behavior is our own is not foolproof. As we saw in the case of bodily self-ascriptions, the self-specifying information that one receives may not be sufficient to settle questions of body ownership, and one must resort to hypotheses about which limb is one's own. On these occasions, Dennett argues that we must perform "experiments" in which we "do something and look to see what moves" (1991, p. 428). Groucho stands before the mirror unsure if what he sees is another or a reflection of himself; to arrive at a conclusion, he moves in certain ways and observes the results. Moreover,

Dennett contends that we apply the same strategy to identify "our own internal states, tendencies, decisions, strengths, and weaknesses" (p. 428). For such internal states, we experiment by constantly telling stories, forever checking and revising them to maximize consistency and inclusiveness among a set of narrative strands.

The pattern of behavior to be brought under a single narrative may be quite complex and puzzling, and this picture of identifying strategies entails the possibility of radical mistakes of identification. Indeed, this arguably occurs. As we saw in the case of thought insertion, a subject concludes that she did not author certain thoughts of which she is introspectively aware. In narrative terms, the center of narrative gravity that organizes most of the narrative strands that make up her autobiography stands at odds with the thought in question. Given how she identifies herself, this thought seems wildly out of character with that identification, so much so that the posited self responsible for her past behaviors simply could not have been the same self that produced it. Being a contentful state, it must have some author; hence she preserves her narrative boundary by assigning that thought to another and contriving a story of insertion to account

for how she came to be aware of it in a direct and unmediated way.

Conflicts in narrative need not be limited to particular thoughts, however. Subjects who suffer from dissociative identity disorder (DID) seem to have more than one self in the same physical body. In these cases, the patterns of speech and behavior admit of no clean integration, and no single agent appears responsible for producing all of them. In fact, bits of behavior may cluster together in ways that encourages positing different selves for different clusters. DID patients often present rather distinct patterns of behavior or "alters"--one may be male and prone to fits of rage, another may talk like a vulnerable young girl, and yet another may be daring and disposed to recklessness, each perhaps calling "himself" or "herself" by different names. When we encounter these subjects, we can't help but think that distinct personalities are in fact taking turns inhabiting and controlling their bodies. Under the narrative picture--and, by extension, NP--this results from the way selves are identified and posited. Such radical ranges of behavior simply offer no one identification of a single self. The best predictive and explanatory theory calls for

attributing particular bits of behavior to separate selves.¹⁴

Dennett does give us a way of understanding how we create selves to interpret patterns of behavior. But one could object that his explanation merely passes the regress problem on to a different level. Suppose that he is correct in thinking that we do find ourselves "engaged in all sorts of behavior" that we then attempt to make cohesive and coherent by positing a narrative center. Can I still be wrong about which narrative is my narrative? Why do I attempt to reconcile just these bits of verbal and nonverbal behavior? Mustn't there also be a reconciler?

Dennett relegates his dismissal of this question to a footnote:

And how do we know that we are doing something? Where do we get the initial bit of self-knowledge we use for this leverage? This has seemed to be an utterly fundamental question to some philosophers (Castañeda 1967, 1968; Lewis 1979; Perry 1979), and has generated a literature of surpassing intricacy. If this is a substantial philosophical problem, there must be something wrong with the 'trivial' answer (but I can't see what): We get our basic, original self-knowledge the same way the

¹⁴ Of course, my discussion of DID only scratches the surface of this fascinating phenomenon. For an interesting and informative exploration of the subject, including several case studies, see Hacking (1995). See also Dennett (1991), Humphrey and Dennett (1989/1998), and Wegner (2002). Wegner likens it to the phenomenon of spirit possession and advances a view quite similar to Dennett's. He argues that we construct virtual selves which are, as the term suggests, not real.

lobster does; we're just wired that way
(1991, p. 428, n2).

This answer hardly satisfies, however. Our respect for our biological boundaries may indeed be due in large part to our "wiring",¹⁵ but it's unclear how the claim survives the analogical leap to the narrative self. Wiring is itself a general term standing for complex biological processes that arise out of the physical connection and interaction of various independent cells. Narrative "cells" share no physical connection, of course, which means that the analogous organizing principle for narrative selves remains mysterious. If the suggestion is instead that we are simply wired to take certain strings of verbal behavior as our own, then the narrative self begins to seem much more like nonfiction than fiction.

Still, we can build on Dennett's provocative idea and in so doing perhaps put the nagging regress worry to rest. The ascriptive regress does not terminate with the ascribing self but instead with ascription. For certain bits of ascriptive behavior produced by minded and language-producing creatures in effect posit certain ascribers for those very acts. Robert Nozick (1981)

¹⁵ The studies regarding body ownership discussed in the previous chapter, however, suggest that this "wiring" doesn't by itself determine one's sense of one's biological boundaries. Self-identifying beliefs play a crucial role as well.

provides an interesting story about how this might work. He argues that we can make sense of the view that no ascribing self exists before an act of nonaccidental self-ascription, that the "I is delineated, is synthesized around the act of reflexive self-referring" (p. 87). Take a self-ascription of the form that one would express as "I am F." The first-person pronoun and its mental analogue function as devices of reflexive self-reference that in effect automatically refer to the entity that tokens them. Accordingly, we can recast them as "The doer of this very act is F." In other words, Nozick suggests that we take Lichtenberg's response to Descartes quite seriously--the self is not revealed in thinking "I am a thinking thing" as Descartes would have it; instead, the grammar or form of the act itself creates a self by positing a thinker as its doer.

How could reflexive self-reference synthesize a self? Nozick argues that reflexive self-reference in general requires no appeal to anything special or mysterious. Take a series of ascriptive acts $A_1, \dots, A_n$, some of which are acts of classification, acts "bringing together things to constitute demarcated entities" (p. 88)--e.g., the thought one would express as "Bill, Dennis, and Tom make up the committee." Now consider a further act of unification and

synthesis A_0 that classifies and demarcates an entity according to $A_1, \dots, A_n$ along with A_0 . The act A_0 classifies all these A -acts--including A_0 itself--as "parts or components or things arising from the same entity E " (p. 89). A_0 posits an entity of which it is a part, and since we've described this possibility without invoking first-person considerations, no further mystery remains. Ascriptive acts involving the first-person pronoun or its mental analogue are merely further instances of reflexive self-reference amenable to this more basic explanation.

Still, we could construe E as the preexisting entity that was the doer of A_0 , but Nozick asks us to take the possibility of self-synthesis as literally as we can:

Can we say (afterwards) that what did A_0 was the entity E which A_0 itself synthesized? Can the rabbit be pulled out of the rabbit? It is some such theory as this that Fichte presents; he speaks of the self as positing itself, also of the self as positing itself as positing itself. ... A_0 seems to be an act of positing the self which, since it reflexively refers to that very act as included within the entity E , posits itself as positing itself, synthesizes itself as synthesizing itself. (p. 89)

The act of self-ascribing unity in a reflexive way thus creates the ascribing self, what Nozick calls (with characteristic wit and flare) a "Fichtetious object" (p. 89).

Nozick is less clear about how these individual acts of synthesis can give rise to the intuitive unity that we associate with selves as we experience them "from the inside." After all, any particular act of reflexive self-reference synthesizes a rather circumscribed object and by itself does not dictate that any particular set of such acts synthesize a single, unified self. Nozick suggests that while any reflexive self-referential act only creates a doer for the very act in question, it sets a precedent for future acts. Precedent doesn't necessitate unity, but it does provide a sort of momentum towards unity, much like an accreting snowball rolling down a hill.¹⁶ Moreover, Nozick argues, much like Dennett, that the self continually rewrites its own history, synthesizing itself so as "to include past entities, including past selves which were synthesized" (p. 91). Ultimately, however, Nozick effectively abandons the unity question by begging it. He effectively packs unity into the act of synthesis itself, concluding that "[t]he I synthesizes itself as having the identity through time of a unified whole" (p. 104).

Though he presents a clever and interesting proposal, Nozick eventually loses faith in it, for it ultimately

¹⁶ We will return to this suggestion and to the question of apparent unity of selves as they appear in Rosenthal (2003) below.

seems to him "too much like a bit of philosophical chicanery, too much froth and too little substance" (p. 114). He finds his synthesized self too insubstantial to bear the necessary philosophical weight. Given his expectations, it's no wonder. He offers the synthesis view as a way to vindicate our intuitions about the puzzling aspects of what he broadly calls "the essence of selves," such as the intuition that a self "stands in a peculiarly intimate relationship to itself," is "conscious of itself as itself" (p. 71), and has special knowledge of itself that can yield self-ascriptions that are IEM (p. 90). Not only must an act of reflexive self-reference synthesize a self, Nozick demands that it must synthesize a self that knows that it refers to itself in this particular way (p. 90). He also wants it to provide at least the basis for a projectible self-conception and the basis of self-care. It's natural that Nozick's view should seem vaporous to him--the questions that he wants his proposal to put to rest can only be suitably answered by a "weighty" proposal.

Still, Nozick's proposal does provide a useful way of characterizing the distinction proponents of IEM want to capture between nonaccidental and accidental self-ascriptions. Non-accidental self-ascriptions are those that involve the first-person pronoun, a device of

reflexive self-reference. Such self-ascriptions are therefore less subject to ascriptive accidents because only acts of self-reflexive reference themselves posit an ascriber of those very acts. Proper names and definite descriptions make no such reflexive reference to their tokeners; they refer in virtue of semantic features independent of their tokenings. Unlike self-reflexive contents, names can be empty and descriptions can go unsatisfied. As a result, ascriptive accidents are more likely to occur.

Moreover, by this point we can see that Nozick's desiderata for a theory of self-ascription no longer remain so desirable. We simply do not have a special sort of awareness of or access to ourselves that needs explaining in terms of a substantial self. Instead, we need an explanation of the appearance of these phenomenon, and an approach like Nozick's can provide a way of understanding this appearance. But such an approach seems dramatically atomistic, and our introspective conscious experience of ourselves is of a unified, persisting, and substantial self. How can we reconcile the two?

Though he devotes his paper "Unity of Conscious and the Self" (2003) to explaining the appearance of conscious unity in terms friendly to his higher-order thought (HOT)

model of consciousness, Rosenthal's approach to the question of the sense of unity we have with regard to our mental lives has much to teach us about our own reconciliation challenge. For it seems to one, from the point of view of consciousness, that all the self-ascriptions that one would normally express with the first-person pronoun refer to the same, single individual. It seems to us that the subject who thought that she was hungry, e.g., later thinks that she had goulash for lunch. For any conscious state that one is in, it seems to one that it participates in a single stream of consciousness, the train of thought of one individual.

To better understand his strategy, we first need at least a general understanding of the HOT model of state consciousness.¹⁷ According to this model, an intentional or sensory state is conscious just in case one has an assertoric thought about it to the effect that oneself is in that state. In other words, for one to be in a conscious state is for one to be conscious of oneself as being in it in a suitable way.¹⁸ If one is in a particular

¹⁷ See Rosenthal (1986b) and (1993) and others reprinted in Rosenthal (forthcoming 2004) for representative discussions of his HOT model.

¹⁸ Rosenthal spells out the qualifier "suitable way" in terms of two caveats about HOTs: they must not arise from any conscious inference and they must represent the states they are about as being

mental state but is in no way aware of being in that state, it will not seem to one from the point of view of consciousness that one is in that state, and hence that state will not be conscious.

Unity becomes a pressing issue for the HOT model since prima facie it offers a fairly atomistic picture of state consciousness. If each conscious state is conscious in virtue of an accompanying HOT that makes one conscious of it as being present, the question arises as to how possessing several distinct HOTs can give rise to a single, unified sense of consciousness. After all, even if each particular HOT represents an individual as being conscious of herself as being in a particular mental state, why think that any two HOTs ascribe states to the same individual?

According to Rosenthal, each HOT "characterizes the self to which it assigns its target solely as the bearer of that target state and, by implication, as the individual that thinks the HOT itself" (2003, p. 30).¹⁹ Though this

present. The former captures the direct and immediate way we seem conscious of our conscious states and the latter is needed to insure that having a higher-order thought about a state makes one conscious of it.

¹⁹ The "by implication" qualifier is quite important. HOTs can't literally be about themselves or all HOTs would be conscious. Since HOTs typically aren't conscious, they must refer to themselves only "by implication." Rosenthal argues that HOTs should not be understood as token reflexives but rather, following Kaplan (1989), as involving a mental analogue of 'I' tantamount to a function from contexts to contents. Just as the first-person pronoun refers to whoever expresses

sparse characterization of the bearer of conscious states does little to unify one's conscious mental states, much like Nozick Rosenthal contends that it does just as little to splinter it:

Nothing in that characterization implies that this bearer is the same from one HOT to the next. But there is also nothing to distinguish one such bearer from any other. And our seeming to be aware in a direct and unmediated way of the self each HOT refers to tilts things towards apparent unity. Since we seem to be directly aware of the self in each case, it seems subjectively as though there is a single self to which all one's HOTs refer, a single bearer for all our conscious states. (p. 332)

One as it were assumes that one is directly and immediately conscious of the same self on each occasion, and that one's conscious thoughts about that self and their expression all have the same referent--barring reason to think otherwise.

We make a similar conservative assumption about the referents of proper names. Rosenthal observes that "we take distinct tokens of a proper name all to refer to the same individual unless something indicates otherwise..." (p. 335). I assume, for example, that all my uses of the name 'Sigmund Freud' refer only to Sigmund Freud. The assumption can prove false--I could discover that no single individual wrote the works or gave the lectures usually

a speech act in which it participates, the mental analogue of 'I' refers to the thinker of the thought in which it figures.

attributed to Freud. Barring strong evidence to the contrary, however, we consider both our own uses of the first-person pronoun and proper names to be referentially constant and consistent.

Nevertheless, Rosenthal agrees that we "identify the individual a first-person thought refers to by appeal to some heterogeneous battery of contingent properties" (2003, p. 345), and this way of identifying ourselves still seems at odds with the way we are conscious of ourselves. For he argues that a modicum of IEM does persist with regard to consciousness. Though one can always be mistaken about who is conscious of being F, Rosenthal contends that mistakes "arguably cannot occur" (p. 340) concerning whether oneself is the individual who is conscious of himself as being F. This is unsurprising given his HOT model. For a state to be conscious--a state of pain, e.g.--simply is for one to be conscious of oneself as being in that state. That is,

the HOT in virtue of which the pain is conscious in effect represents it as belonging to the individual that thinks that HOT. So there is no way to go wrong about whether it's I that I think is in pain. I am conscious of an individual both as being in pain and, in effect, as the one that's conscious of being in pain, and I use the mental analogue of 'I' to refer to that individual. (p. 346)

For a pain to be conscious is for one to have a HOT to the effect that oneself is in pain.

The IEM this affords, argues Rosenthal, does not conflict with the battery proposal. This "thin" immunity only applies to the way in which one's HOTs represent oneself and therefore the way in which one is conscious of oneself. The battery of self-specifying beliefs gives one a way of filling out and distinguishing that conscious self from others. New HOTs add to the battery by enriching the descriptions one has of oneself as a conscious subject. And the unity assumption gives rise to the sense we have that the loose battery of contingent properties picks out a single self.

'Thin', of course, is a relative term, and in the end Rosenthal's thin immunity resembles Shoemaker's "thicker" version in many respects. Rosenthal argues that, unlike Shoemaker, the immunity he favors applies even in nonconscious cases. If one were to see a mirror reflection of oneself limping, one might conclude that one is in a nonconscious state of pain. But Shoemaker might well draw the same conclusion.²⁰ To recognize that it is oneself who one sees limping in the mirror, Shoemaker argues that one

²⁰ Though he would not agree to the possibility of nonconscious pains. See (1996), pp. 226-9. That point of contention is irrelevant in this case.

must have IEM beliefs about oneself that can fund conscious self-identification. Whereas the ascription of pain itself in the mirror case might be open to error--as was Groucho's conclusion regarding Harpo's ruse--IEM applies at the more fundamental level of determining whether oneself is reflected in the mirror. Accordingly, the immunity Rosenthal endorses becomes "thick" very quickly, for it seems that all conscious ascriptions of states and properties--whether the state ascribed is conscious or nonconscious--require that one possess beliefs about oneself that are IEM.

The HOT model also has its own form of self-intimation. Recall that Shoemaker argues that our functional organization is such that if one is in a certain mental state one will be caused to believe that one is in that state. Hence, one cannot have the conscious belief that *p* without also having the belief that one believes that *p*. Similarly, according to the HOT model, a state is conscious if and only if one is conscious of oneself as being in that state in the appropriate way. The "appropriate way" dictates that the HOT in virtue of which a state is conscious must be a first-person thought. As a result, even when a state is conscious but not introspectively conscious, one must have the first-person

belief that oneself is in that state. HOTs need not be conscious--and usually aren't, argues Rosenthal. Nevertheless, his "thin" immunity--like Shoemaker's--results from the intimate connection between any conscious mental state and the IEM self-ascription of the belief that one is in that state.

We should note, however, that Rosenthal's form of self-intimation is only analogous to Shoemaker's. According to the HOT model, consciousness is not an intrinsic property of mental states, and the apparent direct access to our states characteristic of consciousness results not from causal relations between a mental state and its self-ascription but from HOTs that involve no conscious inference. HOTs may indeed be caused by their targets, but Rosenthal contends that HOTs can arise in many different ways. Thus the model leaves room for one to become conscious of another's states, say, if one could token HOTs targeting those states that did not rely upon conscious inference. In this counterintuitive case, one would have a HOT to the effect that "the thinker of this very HOT is in state S" where 'S' picks out the state of another. If the other individual in question were indeed in state S, then it does seem that one would be correct about the state ascribed but mistaken about the identity of

the ascriber--a result clearly proscribed by Shoemaker's thicker IEM.²¹

Still, Shoemaker could argue that this anomalous example instead evinces the inexorable thickness of IEM. For Rosenthal agrees that when a mental state is conscious one cannot be wrong about whether it is oneself who is representing himself as being in that state. So even when one has a noninferential HOT targeting the mental state of another--and it therefore seems to one subjectively that that state is conscious--one ascribes the state to oneself. Hence, Shoemaker might describe the situation differently: One's ascription of the other's state is IEM (in the thick sense) but false because one is not in fact in that state. One has not misidentified oneself but instead was mistaken about which state one was in. Hence, it seems that little separates thick from thin versions of IEM.²²

²¹ This paragraph draws upon Rosenthal's updated and expanded (2003) available on the World Wide Web at <http://web.gc.cuny.edu/cogsci/unity-rev.pdf>.

²² Shoemaker might also argue that if another's states could be conscious for one, then it's no longer clear whether two distinct subjects remain. This response flirts with begging the question, however, for he could not appeal to the premise of weak self-intimation to help draw boundaries for distinct minds. Nevertheless, a response of this kind does seem warranted, for if one could truly be conscious of another's states in the characteristic way that one is conscious of one's own conscious states, one would, in an important sense, be in that state. The HOT responsible for that outcome would then correctly describe oneself. In other words, the individual who was the thinker of that HOT would indeed be the individual who was in the state ascribed by it.

Moreover, under Rosenthal's picture it's unclear what the battery of self-identifying beliefs actually contributes to one's overall sense of mental unity. He does agree that identifying oneself amounts to "saying who it is that one is talking or thinking about when one talks or thinks about oneself" (p. 335). And he also agrees that one appeals to a battery of diverse and contingent properties to identify oneself. However, Rosenthal claims that the battery serves a reference-fixing role as well.

He writes:

For any new first-person thought, the reference that thought makes to oneself is secured by appeal to what other, prior first-person thoughts have referred to, and this process gradually enlarges the stock of self-identifying thoughts available to secure such reference. (p. 335)

But assigning the battery a reference-fixing role conflicts with the Kaplan model Rosenthal favors. According to Kaplan (1989), the first-person pronoun is tantamount to a function from context to contents. The referent of a token of 'I' is determined by its context of utterance (or mental tokening) apart from any ability one might have to identify or to pick out the tokener--reference and identification can come apart. If the battery did secure the reference of the first-person pronoun, mistaken beliefs about oneself

could result in one's tokens of 'I' referring to someone else altogether.

Of course we need not endorse the HOT model of consciousness or follow Rosenthal's construal of the battery to the letter. One crucial lesson to draw from it, however, is that how things seem from the point of view of introspective consciousness may differ radically from how things actually are. When self-ascriptions are introspectively conscious, subjectively it seems to us that we are aware of a robust, unified ascribing subject in a direct and unmediated way. The appearance does need to be explained, but it need not preclude us from giving an account of self-ascription and the ascribing subject in terms friendly to NP. Indeed, the direct and unmediated way in which one is aware of introspected mental states and properties encourages the sense that they all belong to the same individual.

Together, the work of Dennett, Nozick, and Rosenthal suggests a slightly different unity proposal that draws much more heavily upon the battery of identifying beliefs. Not only are we skillful self novelists,²³ we are also accomplished literary critics. Questions do arise as to the authorship of certain narratives, and once authorship

²³ I draw the "novelist" reference from Dennett (1988).

is agreed upon, the characteristics of that narrative form the basis of a profile that can then aid in identifying still other narratives as by that same author. Academic industries have sprung up around whether Shakespeare was indeed the author of all the plays commonly attributed to him (or if in fact they can be attributed to a single author), and certain Edgar Allen Poe scholars have devoted their lives to determining which of many unsigned columns, reviews, and stories were produced by Poe himself.²⁴

On occasion, author and authorship can only be determined together. In the interpretation accompanying David Rosenberg's translation of The Book of J, for example, Harold Bloom argues that a strand of narrative running through the books of Genesis, Exodus, and Numbers were written by a single, brilliant author whom scholars call, simply, J. His strategy is to construct as sharp a picture as possible of J (including her sex) on the basis of her extant writings. To make his case, however, Bloom has to redress what he describes as the "long, sad, enterprise of revising, censoring, and mutilating J" (p.

²⁴ I know a particularly tenacious Poe scholar who even in his retirement is determined to produce the definitive collected works of Poe. Interestingly, he was always quick to offer that though he respected Poe's literary gifts (especially his penchant for coining new words), he wouldn't care to have him as a friend. The point being that by studying his writing (and the many writings about him), this scholar has come to have a fairly rich battery of identifying beliefs about Poe as an author and as a person.

22), which means that he must also argue that certain fragments of narrative currently assigned to a variety of authors are in fact the work of J. His main tool for doing so is his sensitivity to authorial voice, a product of his "experience over half a century as a reader" (p. 21).

In priming the reader for his case, Bloom connects his exercise with more familiar attempts at understanding any author:

This J is my fiction, most biblical scholars will insist, but then each of us carries about a Shakespeare of a Tolstoy or a Freud who is our fiction also. As we read any literary work, we necessarily create a fiction or a metaphor of its author. That author is perhaps our myth, but the experience of literature partly depends upon that myth. (1990, p. 19)

In reading a text, we construct a picture of the author--her literary strengths and weaknesses, favorite themes, habits of mind, expressive powers, etc. To read a narrative is ultimately inseparable from "reading" its author, and to determine the boundaries of a narrative or text we must draw upon the character and qualities of putative authors.

This complementary process is not merely an academic exercise. Consider once again Sacks's surprise at his own reaction. His experience of venomous spite itself says something about the individual responsible for it--perhaps

that he can be petty, envious, and mean. Sacks shifts authorship of this experience to his sickness because he does not believe that he is the type of person who could produce such a reaction. The experience speaks volumes about its author, and Sacks finds that its authorial voice sounds nothing like the one common to the more staid, avuncular thoughts and experiences normally constituting his conscious experience. And if (or when) he does admit that it was not his sickness "talking" but rather himself, he will be more likely to own up to similar experiences in the future. Assigning authorship to a narrative strand amounts to admitting to something about the author, even if it is oneself.

The above remarks form just the foundation of a more complicated proposal, for the ways in which we identify ourselves are due to a host of sources, including the ways in which others identify and describe us.²⁵ Still, the authorship analogy helps to explain how self-ascription draws upon self-identification without assuming a unified, substantive, ascribing self. The battery actively gives rise to a sense of unity by providing a constantly enriching base for interpreting the introspecting subject as well as the states introspected. Conscious self-

²⁵ See, e.g., Neisser (1988).

reflexive self-ascriptions both create and characterize an ascriber, and new conscious thoughts and behaviors that do not conflict with that characterization are incorporated into the battery by default. The battery thus identifies a subject as a rich text would its author. And the richer the battery becomes, the more successful the identification of new bits of narrative behavior.

§4. The Not-So-Naïve-Proposal: A Review

We began in Chapter II with the puzzling relationship between self-identification and self-ascription. Intuition dictates that certain self-ascriptions involve no self-identification whatsoever. We found proof of the power of this intuition in the way in which Chisholm, Wittgenstein, and Anscombe assume--explicitly or implicitly--that certain sorts of ascriptive error are simply impossible. In Wittgenstein's terms, no room for error exists when one self-ascribes properties in a particular way with the first-person pronoun.

Any sufficient account of self-ascription, however, also needs to explain what we called accidental self-ascriptions, or those that do intuitively seem open to self-misidentification. Canonically, these are self-ascriptions that one would express with self-designating

terms other than the first-person pronoun such as proper names, descriptions, and demonstratives. Misidentification results when one is both disposed to assent to what one would express as "a is F" for some self-designating expression 'a' and disposed to deny what one would express as "I am a."

Initially, it appeared that accounts such as Chisholm's that excel at explaining nonaccidental self-ascription had little to say about accidental self-ascription. Similarly, accounts like the broad formulation of NP that take accidental cases as the model for all self-ascriptions seem incapable of capturing our intuitions about nonaccidental ones--particularly first-person ascriptions of mental states and properties. The strengths of one approach highlights the weaknesses of its competitor, and vice versa.

Shoemaker provides a much more compelling and comprehensive account similar to Chisholm's in terms of certain self-ascriptions being IEM. He offers a way to understand both accidental and nonaccidental self-ascriptions in terms of the way in which one is aware of the property or state ascribed. First-person self-ascriptions based upon introspective awareness are nonaccidental because one is aware of the relevant state or

property in a way that is direct and unmediated. One doesn't first observe some individual as, e.g., believing that p and then identify that individual as oneself. According to Shoemaker, one's mental states are self-intimating, and to believe that p is functionally inseparable from the introspective belief one believes that p. In contrast, accidental self-ascriptions are funded by types of awareness that are indirect and mediated. When one observes a reflection in a mirror, for example, one sees that some presented individual is F, but a question can arise as to which individual is the one that is F, including whether that individual is in fact oneself. One must identify that individual by drawing upon antecedent beliefs about oneself or another, and since identification is involved, misidentification (of oneself or another) cannot be precluded.

In Shoemaker's most powerful argument in support of his account, he contends that at least some self-ascriptions must be nonaccidental, for otherwise we would be unable to explain how we are able to ascribe properties to ourselves and to others even in accidental cases. If all self-ascription did depend upon identification--and hence were subject to misidentification--we would fall into a vicious regress. Identifying some presented individual b

as some familiar individual a requires that one already possess beliefs about the properties of a that overlap in some significant measure with the observed properties of b. If one's antecedent beliefs about a themselves depended upon identification, then one must possess antecedent beliefs about the properties of some other familiar individual c that overlap with the observed properties of a, and so on indefinitely. We can halt this regress only by positing self-ascriptions that do not depend upon ascribing any property to oneself and hence are immune to error through misidentification relative to the first-person pronoun. Indeed self-misidentification is possible, Shoemaker argues, because we have beliefs about ourselves that are IEM and mistakenly conclude that some presented individual instantiates those properties we ascribe without first identifying ourselves.

We further observed that Shoemaker's account survives challenges from within the ranks of those who endorse IEM. Evans and Bermúdez argue that certain bodily self-ascriptions are likewise IEM, but empirical investigations into our sense of bodily ownership and the ways in which it can be disrupted indicate otherwise. Though it may appear that we are aware of our bodily properties and limbs in a direct and unmediated way, mistakes in attributions of

ownership can arise. Corporeal beliefs about one's body such as its size, appearance, limb orientation and number, make up one's body image, and by manipulating these beliefs one can be fooled into self-attributing prosthetic limbs or the limbs of another. As a result, however it might appear "from the inside," bodily self-ascriptions are not in fact IEM.

This counterintuitive result provides a model for reconsidering Shoemaker's version of IEM self-ascriptions based upon introspective self-awareness. Here again we observed, with the help of empirical research, that though it seems that one is aware of one's own mental states and properties in a way that is direct and unmediated, we in fact do not have such access. The strong intuition that we cannot make mistakes of self-identification when introspecting results from certain self-ascriptions being introspectively conscious. But mistakes can arise even in these cases, for one can consciously believe oneself to be in certain states that one is not in fact in. The point of view of consciousness is not, as it turns out, a privileged perch from which we witness our own mental lives. IEM proves to be nothing more than a figment of introspective consciousness.

NP, in contrast, accommodates both the intuitions we have about self-ascriptions and the errors to which such thoughts and expressions about ourselves are open. It maps the intuitive difference between nonaccidental and accidental self-ascriptions not in terms of the impossibility of self-misidentification but instead in terms of the amount of self-specifying information brought to bear when ascribing states or properties and the form of the self-ascription. According to NP each of us possesses a battery of self-specifying beliefs, and nonaccidental self-ascriptions--those canonically made with the first-person pronoun--are less likely to involve ascriptive errors because one draws upon the full range of one's beliefs about oneself. Introspection is better thought of as conscious interpretation, the application of a predictive and explanatory strategy useful for making sense of our behavior and experiences. This richness of belief, combined with how the ascribing subject appears from the point of view of introspective consciousness, gives rise to the intuition that one cannot misidentify the subject of first-person self-ascriptions.

NP also puts to rest Shoemaker's regress worries without invoking a special form of self-access. Interpretation often involves interpreting not only which

state one has but whether it is oneself who has it. Accordingly, the self to which we ascribe properties accidentally and nonaccidentally is a narrative self, the theoretical posit specified by one's battery of self-identifying belief. This narrative self arises out of individual narrative ascriptive strands, some of which are self-reflexive and therefore themselves posit an ascriber for that very act. These minimal posits set a kind of precedent for the identity of the ascribing subject, providing the precedent for a narrative self and contributing to a growing sense of conscious unity.

NP has the remarkable consequence--one that many will find hard to accept--that first-person misidentification remains a possibility even with introspectively conscious self-ascriptions. Arguably this does happen in instances of thought insertion. We identify ourselves as particular sorts of thinkers and doers, and when thoughts and actions that we are aware of in a direct and unmediated way run counter to that identification we can go so far as to misidentify the ascriber by attributing those thoughts and actions to another. For one to misidentify oneself is to deny that oneself is the referent of a particular expression that is in fact self-designating. One is introspectively conscious of a property or state when it

seems to one that one is aware of it in a way that is direct and unmediated. But that conscious self-ascription that one would normally express with the first-person pronoun--"I must kill God," e.g.--could dramatically conflict with how one identifies oneself. One could, therefore, deny that one tokened that very thought--"it's the thought of this chap Chris." But to be a competent speaker of English, one must grasp that tokens of the first-person pronoun (and its mental analogue) refer to their tokeners. One thus in effect denies that oneself is the first person posited by the "alien" self-ascription, thereby misidentifying oneself.

In the end, then, contrary to our powerful intuitions, all self-ascription involves identification in some way. Nonaccidental self-ascriptions differ from accidental ones in this respect only in degree and not in kind. We have no special access to our own mental states and properties, and thus we have not been granted immunity against errors of self-identification, even in introspective cases. However disconcerting the persistent possibility of ascriptive error may be, it allows for a comprehensive, unified account of self-ascription that leaves little room for Cartesianism to take root. As a

result, in NP we find a means of demystifying the first person.

§5. Answering the Objections: The Misconception Problem and the Amnesia Problem

Armed with the full version of NP, we are now in a position to return to and to answer the three major objections raised against NP at the outset. First, let's consider the Misconception Problem and the related "Twivienne" puzzle. Recall that this problem questioned whether the battery of self-specifying beliefs could serve to pick oneself out uniquely since one no doubt misidentifies oneself in a host of ways. Moreover, the objection goes, we can imagine a case in which two subjects--Vivienne and Twivienne--possess the same battery of beliefs and hence each identifies herself in the same way. NP seems unable to provide a way of explaining how Vivienne's self-ascriptions ascribe properties to her and not to Twivienne.

NP's answer to this problem should, by now, be clear. First, it confuses reference and identification. One can refer to Freud with the name 'Freud' despite having false beliefs about him. Similarly, one can refer to oneself with names, descriptions, and pronouns even when one has mistaken beliefs about oneself. Just as one's beliefs about Freud don't determine the reference of one's

tokenings of the name 'Freud', one's battery of self-specifying belief doesn't determine one's reference to oneself. Reference is a function of the semantic properties of an expression, and identification is a matter of beliefs--true or false--one has about particular referents. Indeed, instances of misidentification are themselves occasions when reference and identification diverge.

Does this response in effect exacerbate the Twivienne problem? No special difficulty about reference arises, for Vivienne and Twivienne will each succeed in referring to herself uniquely with her tokens of 'I' or its mental analogue. Though it may appear to us as if we have direct and unmediated access to a unique self, we simply do not have the kind of "special privileged access" that Chisholm and Shoemaker would draw upon to settle the dilemma of identification confronting Vivienne and Twivienne. If two identical subjects like Vivienne and Twivienne did truly identify themselves in the same way, the narrative selves they posit to explain and to predict their behavior won't in fact differ. We must accept that our intuitions about identification here simply are not vindicated.²⁶

²⁶ Notice also that according to NP the cottage industry that has developed around the problem of reconciling self-knowledge with

The Amnesia objection, we will recall, questioned the necessity of the battery, for if one could still nonaccidentally self-ascribe properties after having completely lost one's battery of self-specifying belief, not all self-ascriptions involve identification. If anything, this objection now seems even more dramatic in light of our appeal to fictional selves--to lose one's battery of belief is, as it happens, to lose oneself.

This objection requires much care to avoid begging the important questions. When we conjure up examples of amnesiacs to sharpen our intuitions, we tend to envision a subject suffering from what we might call "Hollywood amnesia," where an individual who has forgotten part or all of the catalogue of facts from her past but nevertheless continues to speak her native tongue with perfect fluency.²⁷ In such a case, our amnesiac retains her grasp on the first-person pronoun's communicative function to refer to its token. Accordingly, she will still possess the ability to identify herself as a speaker of English and to interpret herself as the producer of tokens of the

externalism struggles with a misconceived problem. One of its assumptions--namely that one knows what one believes in a direct and privileged way--simply does not survive scrutiny. We in fact do not come to believe what we believe in a way that differs from how we come to believe what another believes--it only appears to us that we do.

²⁷ Indeed, film characters suffering from this type of amnesia often seem to be happier than those around them, as if they had literally forgotten all their cares.

English word 'I'. This also means that she will, upon regaining consciousness following the proverbial sharp blow to the head, immediately begin spinning a new self. She will produce strings of narrative whose spinner will need to be identified, and she will be asked questions about her identity that require answering. She will also have a host of other beliefs to draw upon to fill out this new self-- though she may have forgotten the name of her own mother, she would presumably still remember, e.g., that a mother is one's female parent, and that a parent is one's forebear and so must be older than one, that a forebear is ..., and so on. If our amnesiac does eventually discover her forgotten past, she links the pre- and post-accident selves by positing a single unifying narrative center that explains the entire range of data.

If, however, her amnesia is so severe that she has indeed forgotten how to speak, intuition ceases to be instructive. A proponent of NP would indeed be committed to the claim that such a subject could not self-ascribe properties and states at all. That seems to me to be a reasonable conclusion, perhaps the only more reasonable one being that we simply do not know what to say in such a case. Clearly those who argue that this severe amnesiac could still make self-ascriptions cannot simply assert that

she does so in virtue of retaining a fundamental ability to self-ascribe properties without begging the question outright.

§6. The Asymmetry Problem

Last, we turn to the Asymmetry Problem. This problem arose out of NP's denial of a difference in kind between accidental and nonaccidental self-ascriptions.

Nonaccidental self-ascriptions do not seem to be made on the basis of evidence whereas accidental self-ascriptions and the ascriptions of properties to others do. Indeed, this seems to be an obvious feature of self-ascription that any adequate theory must respect and explain. NP appears to do just the opposite, however, for under it the apparent iron-clad security against self-misidentification that nonaccidental ascriptions intuitively enjoy results from the richness of one's battery of self-identificatory beliefs. But this is to say that nonaccidental self-ascriptions are made on the basis of more evidence than accidental ones.

NP does dictate that identificatory evidence plays a key role in one's nonaccidental self-ascriptions. But here again we need to distinguish how things appear to us and how they are. Most beliefs we have about ourselves are

most likely normally not conscious. When we identify others (and on occasion ourselves) for self-ascription we often engage in conscious inferences that bring to bear beliefs about those individuals. This does not mean, however, that self-ascriptions that do not involve conscious inference involve no inference whatsoever. As we saw in several instances above, empirical research strongly suggests that we often make nonconscious inferences about limb and thought ownership. Moreover, the stark contrast between the evidence brought to bear in the first-person and third-person cases oversimplifies the matter. When one ascribes pain to a close friend or family member, it doesn't usually seem to one that one sees someone in pain and then identifies that person as, say, one's brother. One simply sees that one's brother is in pain.²⁸ No doubt one has many more identificatory beliefs about one's brother than one's acquaintances, so we should expect--and NP accommodates--a similar asymmetry in this case. The asymmetry arises because we don't usually consciously appeal to the wealth of evidence we have about ourselves and those most familiar to us. As a result, questions

²⁸ Cf. Wittgenstein's remark that one is not of the opinion that another has a soul (1970, p. 178). That is, barring special circumstances one doesn't consciously arrive at the conclusion that someone is a living human being. In the usual case one simply acts toward another as if she has a soul.

regarding whether oneself--or in some cases another--is the bearer of the property or state seem out of place.

§7. Conclusion: Avenues of Exploration

Over the course of this extended discussion of identification and its role in self-ascription we have deepened our understanding of the ways in which we ascribe properties to ourselves. Further understanding usually begets further questions, however, and this case offers no exception. I therefore want to conclude by considering a few outstanding questions and avenues of further exploration encouraged by the naïve proposal of self-ascription endorsed and defended here.

We have focused mainly on identification as it relates to language users, but studies in comparative psychology strongly suggest that creatures other than ourselves recognize themselves in ways importantly similar to ours. Lobsters may not eat themselves when hungry, but chimpanzees seem to exhibit a higher level of self-identification. They eventually respond to mirrors with what very much seems like spontaneous or unconditioned self-exploratory behavior, apparently using their reflection to examine parts of their bodies that they cannot see directly. Moreover, research by Gordon Gallup

(1970) and others shows that when a small mark is surreptitiously applied to their foreheads, they touch the mark significantly more often in the presence of a mirror than in its absence. Are chimpanzees identifying themselves? If so, how? What cognitive or narrative resources do they bring to bear to identify themselves that animals who don't pass the "mark test" lack? Gallup's explanation for various species' (including human infants) success at mirror self-recognition relies upon some of the same premises and assumptions appealed to by Shoemaker to argue for IEM. Our understanding of the latter might very well extend our understanding of the former.

Much work also needs to be done regarding various psychopathologies and self-identification. NP explains thought insertion in terms of first-person misidentification--a subject has an experience so anomalous that she interprets it as totally at odds with the way in which she identifies herself, ultimately attributing it to another. Stephens and Graham (2000) argue that this explanation is the best one, but that we also need to distinguish between merely having a thought and thinking it. That is, these agents lack the appropriate sense of agency towards their thoughts while still preserving some sense of subjective ownership. As a result, they contend

that these attributions of subjective ownership remain IEM. However, if what we've said about NP is correct, the sense of subjective ownership is itself a product of identification and so attributions of thought ownership are also not IEM. We need a better account of the relationship between the sense of agency and the sense of subjective ownership with regard to thoughts. One could perhaps argue, armed with NP, that the sense of subjective ownership is dependent in large part upon the sense of agency. The success of this strategy remains to be seen.

Similarly, the battery's significance for one's sense of bodily ownership needs to be fleshed out in much more detail. Participants in this literature often assume that states of bodily awareness must either be subpersonal or conscious. However, as we have seen, self-specifying beliefs need not be conscious, and nonconscious beliefs play an important role in dictating the way our bodies appear to be ours in a direct and unmediated way. Beliefs about agency and oneself as a causal force also no doubt contribute to the sense of ownership. The challenge is to spell out in detail how the perceptual systems that provide self-specifying information relate and inform the ways in which we identify ourselves and how those ways give rise to a conscious sense of ownership.

These are formidable but exciting and worthwhile challenges to pursue, and I look forward to pursuing them. Interestingly, to pursue them is, in a sense, to pursue oneself. In her appropriately titled book None to Accompany Me, Nadine Gordimer perhaps locates a persistent truth when she writes: "Everyone ends up moving alone towards the self" (1994, p. 306). I can think of no more inviting destination.

Bibliography

- Adler, Jonathan (2002). Belief's Own Ethics. Cambridge: MIT Press.
- Albritton, Rogers (1995). "Comments on 'Moore's Paradox and Self-Knowledge'." Philosophical Studies 77, pp. 229-239.
- Almog, Joseph; Perry, John; Wettstein, Howard (Eds.) (1989). Themes from Kaplan. Oxford: Oxford University Press.
- Anderson, James R. (1984). "The Development of Self-Recognition: A Review." Developmental Psychobiology 17, 1, pp. 35-49.
- Anderson, James R. (1986). "Mirror-Mediated Finding of Hidden Food by Monkeys (Macaca tonkeana and M. fascicularis)." Journal of Comparative Psychology 100, 3, pp. 237-242.
- Anderson, James R. (1994). "The Monkey in the Mirror: A Strange Conspecific." In Parker et al, pp. 315-329.
- Anderson, James R. (2001). "Self and Others in Nonhuman Primates: A Question of Perspective." Psychologia 44, pp. 3-16.
- Anscombe, G. E. M. (1975). "The First Person." In Guttenplan, pp. 45-65.
- Armstrong, David M. (1963). "Is Introspective Knowledge Incorrigible." The Philosophical Review 72, 4, pp. 417-432.
- Armstrong, David M. (1968). A Materialist Theory of the Mind. London: Routledge & Kegan Paul.
- Austin, J. L. (1962). How to Do Things with Words. Cambridge: Harvard University Press.
- Ayer, A. J. (1946). Language, Truth and Logic. New York: Dover.
- Bahrick, Lorraine E., and Watson, J. S. (1985). "Detection of Intermodal Proprioceptive-Visual Contingency as a

- Potential Basis of Self-Perception in Infancy." Developmental Psychology 32, pp. 963-973.
- Bahrnick, Lorraine E., Moss, Lisa, and Fadil, Christine (1996). "Development of Visual Self-recognition in Infancy." Ecological Psychology 8, pp. 189-208.
- Bauer, Russell M. (1984). "Autonomic Recognition of Names and Faces: A Neuropsychological Application of the Guilty Knowledge Test." Neuropsychologia 22, pp. 457-469.
- Bauer, Russell M. (1986). "The Cognitive Psychophysiology of Prosopagnosia." In Aspects of Face Processing, ed. H. Ellis, M. Jeeves, F. Newcombe, and A. Young. Martinus Hijhoff, Dordrecht.
- Bekoff, Marc; Allen, Colin; Burghardt, Gordon M., eds. (2002). The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition. Cambridge: MIT Press.
- Beirce, Ambrose (1911). The Devil's Dictionary. New York: A. & C. Boni.
- Beis, J. M., André, J. M., Barre, A., Paysant, J. (2001). "Mirror Images and Unilateral Spatial Neglect." Neuropsychologia 39, pp. 1444-1450.
- Bermúdez, José Luis; Marcel, Anthony; Eilan, Naomi, eds. (1995). The Body and the Self. Cambridge: MIT Press.
- Bermúdez, José Luis (1995a). "Aspects of the Self." Inquiry 38, pp. 489-501.
- Bermúdez, José Luis (1995b). "Transcendental Arguments and Psychology." Metaphilosophy 26, pp. 379-401.
- Bermúdez, José Luis (1995c). "Ecological Perception and the Notion of a Nonconceptual Point of View." In Bermúdez et al (1995), pp. 153-174.
- Bermúdez, José Luis (1998). The Paradox of Self-Consciousness. Cambridge: MIT press.
- Boër, Steven E. and Lycan, William G. (1980). "Who, Me?" The Philosophical Review 89, 3, pp. 427-466.

- Boër, Steven E. and Lycan, William G. (1986). Knowing Who. Cambridge: MIT Press.
- The Book of J (1990). (David Rosenberg Trans.). Interpretation by Harold Bloom. New York: Vintage Books.
- Borges, Jorge Luis (1998). Collected Fictions. New York: Viking Penguin.
- Borges, Jorge Luis (1999). Selected Poems. New York: Viking Penguin.
- Bottini, Gabriella, Bisiach, Edoardo, Sterzi, Roberto, and Vallar, Giuseppe (2002). "Feeling Touches in Someone Else's Hand." Neuroreport 13, 2, pp. 249-252.
- Botvinick, Matthew, and Cohen, Jonathan (1998). "Rubber Hands 'Feel' Touch That Eyes See." Nature 391, pp. 756.
- Breen, N., Caine, D., Coltheart, M., Hendy, J., and Roberts, C. (2000). "Towards and Understanding of Delusions of Misidentification: Four Case Studies." Mind & Language 15, 1, pp. 74-110.
- Brewer, Bill (1992). "Self Location and Agency." Mind 101, pp. 17-34.
- Brewer, Bill (1995). "Bodily Awareness and the Self." In Bermúdez et al (1995), pp. 291-310.
- Butterworth, George (1995). "An Ecological Perspective on the Origins of the Self." In Bermúdez et al (1995), pp. 87-106
- Campbell, John (1994). Past, Space, and Self. Cambridge: MIT Press.
- Campbell, John (1995). "The Body Image and Self-Consciousness." In Bermúdez et al (1995), pp. 29-42.
- Campbell, John (1999a). "Immunity to Error Through Misidentification and the Meaning of a Referring Term," Philosophical Topics 26, pp. 89-104.
- Campbell, John (1999b). "Schizophrenia, The Space of Reasons, and Thinking as a Motor Process." The Monist 82, 4, pp. 609-625.

- Capgras, J. and Reboul-Lachaux, J. (1923). L'illusion des 'Sosies' dans un Délire Systématisé Chronique. Bulletin de la Société Clinique de Médecine Mentale, 2, pp. 6-16.
- Cassam, Quassim, ed. (1994). Self-Knowledge. New York: Oxford University Press.
- Cassam, Quassim (1995). "Introspection and Bodily Self-Ascription." In Bermúdez et al (1995), pp. 311-336.
- Cassam, Quassim (1997). Self and World. New York: Oxford University Press/Clarendon Press.
- Castañeda, Hector-Neri (1966). "'He': A Study in the Logic of Self-Consciousness." Ratio 8, pp. 130-157.
- Castañeda, Hector-Neri (1968). "On the Logic of Attributions of Self-Knowledge to Others." Journal of Philosophy 65, 15, pp. 439-456.
- Castañeda, Hector-Neri (1969). "On the Phenomeno-Logic of the I." Proceedings of the XIV Internationalen Congress of Philosophy, 3, pp. 260-266.
- Castañeda, Hector-Neri (1990). "Indexicality: The Transparent Subjective Mechanism for Encountering a World." Noûs 24, pp. 735-749.
- Cavell, Stanley (1976). Must We Mean What We Say?. Cambridge: Cambridge University Press.
- Chapman, J. (1996). "Early Symptoms of Schizophrenia." British Journal of Psychiatry 12, pp. 225-251.
- Chapman, L. J. and Chapman, J. P. (1988). "The Genesis of Delusions." In Delusional Beliefs, ed. T. Oltmanns and B. Maher. Wiley.
- Chisholm, Roderick (1994). "On the Observability of the Self." In Cassam (1994), pp. 94-108.
- Chisholm, Roderick (1981). The First Person: An Essay on Reference and Intentionality. Minneapolis: University of Minnesota Press, and Hassocks: The Harvester Press.

- Christodoulou, G. N. ed. (1984). Delusional Misidentification Syndromes. Basil, Switzerland: Karger.
- Cole, Jonathan, and Paillard, Jacques (1995). "Living Without Touch and Peripheral Information about Body Position and Movement: Studies with Deafferented Subjects." In Bermúdez, Marcel, and Eilan (1995), pp. 245-266.
- Cole, Jonathan, Sacks, Oliver, and Waterman, Ian (2000). "On the Immunity Principle: A View from a Robot." Trends in Cognitive Sciences 4, 5, p. 167.
- Corazza, E., Fish, W., and Gorvett, J. (2002). Who is I? Philosophical Studies 107, pp. 1-21.
- Cotard, J. (1882). "Du Délire des Négations." Archives de Neurologie 4, pp. 152-170, 282-295.
- Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., Jeannerod, M. (1997). "Looking for the Agent: An Investigation Into Consciousness of Action and Self-Consciousness in Schizophrenic Patients." Cognition 65, pp. 71-86.
- Davidson, Donald (2001). Subjective, Intersubjective, Objective. Oxford: Oxford University Press.
- Davies, Martin and Coltheart, Max (2000). "Introduction: Pathologies of Belief." Mind & Language 15, 1, pp. 1-46.
- Dennett, Daniel C. (1978). "Where Am I?" In Dennett, Brainstorms. Montgomery, VT: Bradford Books, pp. 310-323.
- Dennett, Daniel C. (1987). The Intentional Stance. Cambridge: The MIT Press.
- Dennett, Daniel C. (1988). "Why Everyone Is a Novelist." Times Literary Supplement September, pp. 16-22.
- Dennett, Daniel C. (1991). Consciousness Explained. Boston: Little, Brown and Company.
- Dennett, Daniel C. (1992). "The Self as Center of Narrative Gravity." In F. Kessel, P. Cole and D.

- Johnson, eds, Self and Consciousness: Multiple Perspectives, Hillsdale, NJ: Erlbaum, pp. 103-115.
- Dennett, Daniel C. (1993). "The Message Is: There Is no Medium." Philosophy and Phenomenological Research 53, 4 (December), pp. 889-931.
- Dennett, Daniel C. (1998). Brainchildren: Essays on Designing Minds. Cambridge: The MIT Press.
- Dennett, Daniel C. (1996). Kinds of Minds. New York: Basic Books.
- Descartes, René (1985). The Philosophical Writings of Descartes, Volumes 1 and 2, trans. by J. Cottingham et al. Cambridge: Cambridge University Press.
- Devitt, Michael (1984). "Thoughts and Their Ascription." Midwest Studies in Philosophy IX, pp. 385-420.
- Dixon, J.C. (1957). "Development of Self Recognition." Journal of Genetic Psychology 91, pp. 251-256.
- Donnellan, K. (1966). "Reference and Definite Descriptions." Philosophical Review 75, pp. 281-304.
- Drestke, Fred (1990). "Misrepresentation." In Lycan (1990).
- Dretske, Fred (1994/5). "Introspection." Proceedings of the Aristotelian Society 65, pp. 263-78.
- Dretske, Fred (1997). "The Mind's Awareness of Itself." Philosophical Studies 95, 1-2, pp. 103-124.
- Eilan, N., McCarthy, R., and Brewer, M. W. eds. (1993). Spatial Representation: Problems in Philosophy and Psychology. Oxford: Basil Blackwell.
- Eilan, Naomi, Marcel, Anthony, Bermúdez, José Luis (1995). "Self-Consciousness and the Body: An Interdisciplinary Introduction." In Bermúdez et al (1995), pp. 1-28.
- Eilan, Naomi (1995). "The First Person Perspective." Proceedings of the Aristotelian Society 95, pp. 51-66.
- Eilan, Naomi (2000). "On Understanding Schizophrenia." In D. Zahavi and J. Parnas (Eds.) Exploring the Self:

Philosophical and Psychopathological Perspectives on Self Experience. Amsterdam: John Benjamins, pp. .

- Ellis, J. D., Whitley, J. and Luauté, J. P. (1994).
"Delusional Misidentification: The Three Original Papers on the Capgras, Fregoli and Intermetamorphosis Delusions." History of Psychiatry 5, pp. 117-146.
- Ellis, H. D., Young, A. W., Quayle, A. H., and De Pauw, K. W. (1997). "Reduced Autonomic Responses to Faces in Capgras Delusion." Proceedings of the Royal Society: Biological Sciences B264, pp. 1085-1092.
- Ellis, H. D. and Young, A. W. (1990). "Accounting for Delusional Misidentifications." British Journal of Psychiatry 157, pp. 239-248. Reprinted in Young 1998.
- Epstein, R., Lanza, R. P., and Skinner, B. F. (1981).
"'Self-Awareness' in the Pigeon." Science 212, pp. 695-696.
- Evans, Gareth (1982). The Varieties of Reference. Oxford: Oxford University Press.
- Evans, Gareth (1985). Collected Papers. New York: Oxford University Press/Clarendon Press.
- Farné, A., Pavani, F., Meneghello, F., and Ladavas, E. (2000). "Left tactile Extinction Following Visual Stimulation of a Rubber Hand." Brain 123, pp. 2350-2360.
- Ferrari, Michel, and Sternberg, Robert J. eds. (1998).
Self-Awareness: Its Nature and Development. New York: The Guilford Press.
- Fish, F. J. (1962). Fish's Schizophrenia third edition, ed. M. Hamilton. Wright, 1984.
- Fish, F. J. (1985). Clinical Psychopathology: Signs and Symptoms in Psychiatry, M. Hamilton (ed.). Wright.
- Flanagan, Owen (1992). Consciousness Reconsidered. MIT Press.
- Flanagan, Owen (1994). "Multiple Identity, Character Transformation, and Self-Reclamation." In Philosophical Psychopathology, G. Graham and G. L. Stephens (eds.). Cambridge: MIT Press.

- Fodor, J. A. (2003). "Having Concepts; A Brief Refutation of the 20th Century." Keynote speech, Sixth Annual Graduate Student Philosophy Conference, March 15, 2003. New York, NY.
- Foley, J. M. and Breslau, L. (1982). A New Syndrome of Delusional Misidentification. Annals of Neurology 12, p. 76.
- Frege, Gottlob (1892/1997). "On Sinn and Bedeutung." Max Black trans. Reprinted in The Frege Reader, Michael Beaney ed. Oxford: Blackwell Publishers, pp. 172-180.
- Frege, Gottlob (1918/1997). "Thought." Peter Geach and R. Stoothoff trans. Reprinted in The Frege Reader, Michael Beaney ed. Oxford: Blackwell Publishers, pp. 172-180.
- Freud, Sigmund (1966). Introductory Lectures on Psychoanalysis. James Strachey (Trans.). New York: W. W. Norton & Company.
- Frith, Christopher D. (1979). "Consciousness, Information Processing, and Schizophrenia." British Journal of Psychology 1324, pp. 225-235.
- Frith, Christopher D. (1987). "The Positive and Negative Symptoms of Schizophrenia." Psychological Medicine 17, pp. 631-648.
- Frith, Christopher D. (1992). The Cognitive Neuropsychology of Schizophrenia. New Jersey: Lawrence Erlbaum Associates.
- Frith, Christopher D. (1995). "Consciousness is for Other People." Behavioral and Brain Sciences 18, pp. 682-683.
- Frith, Christopher D. (2002). "Attention to Action and Awareness of Other Minds." Consciousness and Cognition 11, pp. 481-487.
- Gallagher, Shaun (1995). "Body Schema and Intentionality." In Bermúdez, Marcel, and Eilan, pp. 225-244).

- Gallagher, Shaun (2000). "Self-Reference and Schizophrenia: A Cognitive Model of Immunity to Error through Misidentification." In Zahavi (2000).
- Gallagher, Shaun and Marcel, Anthony J. (1999). "The Self in Contextualized Action." Journal of Consciousness Studies, 6, 4, pp. 4-30.
- Gallagher, Shaun (2000). "Reply to Cole, Sacks, and Waterman." Trends in Cognitive Sciences, 4, 5, pp. 167-168.
- Gallup, Gordon G., Jr. (1970). "Chimpanzees: Self-Recognition." Science 167, pp. 86-87.
- Gallup, Gordon G., Jr. (1982). "Self-Awareness and the Emergence of Mind in Primates." American Journal of Primatology 2, pp. 237-248.
- Gallup, Gordon G., Jr. (1983). "Toward a Comparative Psychology of Mind." In Mellgren (ed.), pp. 473-510.
- Gallup, Gordon G., Jr. (1987). "Self-Awareness." In Mitchell, G. and Erwin, J. (eds), pp. 3-16.
- Gallup, Gordon G., Jr. (1994). "Self-Recognition: Research Strategies and Experimental Design." In Parker et al, pp. 35-50.
- Gallup, Gordon G., Jr. (1997). "On the Rise and Fall of Self-Conception in Primates." In Snodgrass and Thompson (eds.), pp. 73-82.
- Gallup, Gordon G., Jr., and Povinelli, Daniel J. (1993). "Mirror, Mirror on the Wall Which is the Most Heuristic Theory of Them All? A Response to Mitchell." New Ideas in Psychology 11, 3, pp. 327-335.
- Gallup, G.G., Jr., Povinelli, D. J., Suarez, S.D., Anderson, J.R., Lethmate, J., and Menzel, E.-W. J. (1995). "Further Reflections on Self-Recognition in Primates." Animal Behaviour 50, pp. 1525-1532.
- Gallup, Gordon G., Jr., Anderson, James R., Shillito, Daniel J. (2002). "The Mirror Test." In Bekoff et al (2002), pp. 325-333.

- Garrett, Brian (1995). "Wittgenstein and the First Person." Australasian Journal of Philosophy 73, 3, pp. 347-355.
- Geach, P. T. (1957). Mental Acts. New York: Humanities Press, Inc.
- Geach, P. T. (1957/1972). "On Beliefs about Oneself," in Logic Matters. Oxford: Basil Blackwell, pp. 128-129. Originally in Analysis, XVIII, 1 (October), pp. 23-24.
- Gerrans, Philip (1999). "Delusional Misidentification as Subpersonal Disintegration." The Monist, vol. 82, no. 4, pp. 590-608.
- Gerrans, Philip (2000). "Refining the Explanation of Cotard's Delusion." Mind & Language, vol. 15, no. 1, pp. 111-122.
- Georgieff, Nicholas and Jeannerod, Marc (1998). "Beyond Consciousness of External Reality: A 'Who' System for Consciousness of Action and Self-Consciousness." Consciousness and Cognition 7, pp. 465-477.
- Gillett, G. (1986). "Multiple Personality and the Concept of the Person." New Ideas in Psychology 4, pp. 173-184.
- Gordimer, Nadine (1994). None to Accompany Me. New York: Farrar, Straus and Giroux.
- Graham, George (1999a). "Self-Consciousness, Psychopathology, and Realism About Self." Anthropology and Philosophy 3, pp. .
- Graham, George (1999b). "Fuzzy Fault Lines: Selves in Multiple Personality Disorder." Philosophical Explorations 3, pp. 159-174.
- Graham, George and Stephens, G. Lynn (1994). "An Introduction to Philosophical Psychopathology: Its Nature, Scope and Emergence." In Philosophical Psychopathology, ed. G. Graham and G. Stephens. Cambridge: MIT Press.
- Greenwood, John D. (1991). "Self-Knowledge: Looking in the Wrong Direction." Behavior and Philosophy 19, 2, pp. 35-47.

- Gunderson, Keith (1970). "Asymmetries and Mind-Body Perplexities." In Michael Radner and Stephen Winokur (Eds.) Minnesota Studies in the Philosophy of Science, IV. Minneapolis: University of Minnesota Press, pp. 273-309.
- Guttenplan, Samuel (Ed.) (1975). Mind and Language: Wolfson College Lectures, 1974. Oxford: Oxford Clarendon Press.
- Hacking, Ian (1995). Rewriting the Soul: Multiple Personality and the Sciences of Memory. Princeton University Press.
- Hauser, M.D., Kralik, J., Botto-Mahan, C., Garrett, M., & Oser, J. (1995). "Self-Recognition in Primates: Phylogeny and the Salience of Species-Typical Features." Proceedings of the National Academy of Sciences, USA 92, pp. 10811-10814.
- Henrich, D. (1971). "Self-Consciousness, a Critical Introduction to a Theory." Man and World 4, pp. 3-28.
- Hume, David (1978). A Treatise of Human Nature. L. A. Selby-Bigge and P. H. Nidditch, eds. Second Edition. Oxford: Oxford University Press. Originally published, 1739-1740.
- Humphrey, Nicholas and Dennett, Daniel C. (1989/1998). "Speaking for Our Selves." In Dennett, Brainchildren: Essays on Designing Minds. Cambridge: MIT Press, pp. 31-58. Originally in Raritan: A Quarterly Review, IX (1), Summer 1989, pp. 68-98.
- Humphrey, Nicholas (1978/1983). "Nature's Psychologists." New Scientist 29 (June 1978), pp. 900-904. Reprinted in Consciousness Regained. Oxford: Oxford University Press.
- Hurley, Susan L. (1998). Consciousness in Action. Cambridge: Harvard University Press.
- Hyatt, Charles W. and Hopkins, William D. (1994). "Self-Awareness in Bonobos and Chimpanzees: A Comparative Perspective." In Parker et al. 1994, pp. 248-253.
- Ishiguro, Hidé (1980). "The Primitiveness of the Concept of a Person." In Van Straaten (1980).

- Kaney, S. and Bentall, R. P. (1989). "Persecutory Delusions and the Self-Serving Bias." Journal of Nervous and Mental Disease 180, pp. 773-780.
- Kant, Immanuel (1950). Prolegomena to Any Future Metaphysics. Trans. Mahaffry-Carus. New York: The Liberal Arts Press.
- Kant, Immanuel (1965). Critique of Pure Reason. Trans. Norman Kemp Smith. New York: St. Martin's Press.
- Kaplan, David (1989). "Demonstratives." In Joseph Almog, John Perry, Howard Wettstein.
- Katz, Jerrold J. (1977). Propositional Structure and Illocutionary Force: A Study of the Contribution of Sentence Meaning to Speech Acts. New York: Thomas Y. Crowell Company, Inc.
- Katz, Jerrold J. (1988). Cogitations. Oxford: Oxford University Press.
- Keenan, Julian Paul, McCutcheon, Bruce, Freund, Stefanie, Gallup Jr., Gordon G., Sanders, Glenn, Pascual-Leone, Alvaro (1999). "Left Hand Advantage in a Self-Face Recognition Task." Neuropsychologia 37, pp. 1421-1425.
- Keenan, J.P., Nelson, A., Nelson, M., Nelson, O.C., Pascual-Leone, A. (2001). "Self-Recognition and the Right Hemisphere." Nature 409, p. 305.
- Keenan, Julian Paul, Freund, Stefanie, Hamilton, Roy H., Ganis, Giorgio, Pascual-Leone, Alvaro (2000). "Hand Response Differences in a Self-Face Identification Task." Neuropsychologia 38, pp. 1047-1053.
- Kenny, Anthony (1984). The Legacy of Wittgenstein. Oxford: Basil Blackwell.
- Kinderman, P. (1994). "Attentional Bias, Persecutory Delusions and the Self Concept." British Journal of Medical Psychology 67, pp. 53-66.
- Kinsbourne, Marcel (1995). "Awareness of One's Own Body: An Attentional Theory of Its Nature, Development, and Brain Basis." In Bermúdez et al (1995), pp. 205-223.

- Knoblich, Günther (2002). "Self-Recognition: Body and Action." Trends in Cognitive Sciences 6, 11, pp. 447-449.
- Kripke, Saul (1980). Naming and Necessity. Cambridge: Harvard University Press.
- Lewis, David (1979). "Attitudes De Dicto and De Se." The Philosophical Review 88, pp. 513-543; reprinted in David Lewis (1983) Philosophical Papers: Volume 1. Oxford: Oxford University Press, pp. 133-155.
- Lewis, M. and Brooks-Gunn, J. (1979). Social Cognition and the Acquisition of Self. New York: Plenum Press.
- Lycan, William (1990). Mind and Cognition. ?
- Lycan, William (2001). "Perspectival Representation and the Knowledge Argument." Analysis ??
- Mach, Ernst (1914). The Analysis of Sensations. C. M. Williams and Sydney Waterlow (Trans.). Chicago: Open Court.
- Mackie, J. L. (1980). "The Transcendental 'I'." In Van Straaten (1980).
- Maher, Brendan A. (1974). "Delusional Thinking and Perceptual Disorder." Journal of Individual Psychology 30, pp. 98-113.
- Maher, Brendan A. (1988a). "Anomalous Experience and Delusional Thinking: The Logic of Explanations." In T. F. Oltmanns and B. A. Maher (Ed.) Delusional Beliefs. New York: Wiley Interscience.
- Maher, Brendan A. (1988b). "Delusions as the Product of Normal Cognitions." In T. F. Oltmanns and B. A. Maher (Ed.) Delusional Beliefs. New York: Wiley Interscience.
- Maher, Brendan A. (1992a). "Models and Methods for the Study of Reasoning in Delusions." European Review of Applied Psychology 42, pp. 97-102.
- Maher, Brendan A. (1992b). "Delusions: Contemporary Etiological Hypotheses." Psychiatric Annals 22, pp. 260-268.

- Maher, Brendan A. (1999). Anomalous Experience in Everyday Life: Its Significance for Psychopathology. The Monist, vol. 82, no. 4, pp. 547-570.
- Maher, B. A. and Ross, J. (1984). "Delusions." In H. Adams and P. Sutker (Eds.) Comprehensive Handbook of Psychopathology. New York: Plenum.
- Marcel, Anthony J. (1983a). "Conscious and Unconscious Perception: Experiments on Visual Masking and Word Recognition." Cognitive Psychology, 15, pp. 197-237.
- Marcel, Anthony J. (1983b). "Conscious and Unconscious Perception: An Approach to the Relations Between Phenomenal Experience and Perceptual Processes." Cognitive Psychology, 15, pp. 238-300.
- Marcel, Anthony J. (1993). "Slippage in the Unity of Consciousness." In Experimental and Theoretical Studies of Consciousness, Ciba Foundation Symposium No. 174. Chichester: Wiley.
- Marshall, J. C. (1996). "Do I Really Look Like That?" Cognitive Neuropsychiatry, vol. 1, no. 2, pp. .
- McGinn, Colin (1983). The Subjective View. Oxford: Oxford University Press.
- McDonald, C., Smith, B., and Wright, C. eds. (1998). Knowing Our Own Minds: Essays on Self-Knowledge. Oxford: Oxford University Press.
- McDowell, J. (1994). Mind and World. Cambridge: Harvard University Press.
- Mellgren, R.L. (ed.) (1983). Animal Cognition and Behavior. Amsterdam: North-Holland Publishers.
- Mellor, C. H. (1970). "First Rank Symptoms of Schizophrenia." British Journal of Psychiatry 117, pp. 15-23.
- Meeks, Roblin R. (2001). "How Primitive is Self-Consciousness?: Immunity to Error Through Misidentification and Autonomous Nonconceptual Content." In Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society. Johanna

- D. Moore and Keith Stenning (Eds.). London: Lawrence Erlbaum Associates, pp. 628-632.
- Miller, William Ian (1997). The Anatomy of Disgust. Cambridge: Harvard University Press.
- Millikan, R. G. (1990). "The Myth of the Essential Indexical." Noûs 24, pp. 723-734.
- Mitchell, G. and Erwin, J. (eds.) (1987). Comparative Primate Biology, Vol. 2, Part B: Behavior, Cognition, and Motivation. New York: Alan R. Liss, Inc.
- Mitchell, Robert W. (1993). "Recognizing One's Self in a Mirror? A Reply to Gallup and Povinelli, De Lannoy, Anderson, and Byrne." New Ideas in Psychology 11, 3, pp. 351-377.
- Mitchell, Robert W. (1994). "Multiplicities of Self." In Parker et al, pp. 81-107.
- Mitchell, Robert W. (1997a). "A Comparison of the Self-awareness and Kinesthetic-Visual Matching Theories of Self-Recognition: Autistic Children and Others." In Snodgrass, J.G. and Thompson, R.L. (eds.), pp. 39-62.
- Mitchell, Robert W. (1997b). "Kinesthetic-Visual Matching and the Self-Concept as Explanations of Mirror-Self-Recognition." Journal for the Theory of Social Behavior 27, 1, pp. 17-39.
- Mitchell, Robert W. (2002). "Kinesthetic-Visual Matching, Imitation, and Self-Recognition." In Bekoff et al (2002), pp. 345-351.
- Mollon, P. (1987). "Self-awareness, Self-Consciousness, and Preoccupation with Self." In K. Yardley & T. Honess, eds. Self and Identity: Psychosocial Perspectives. Wiley.
- Moore, G. E. (1942). "A Reply to My Critics." In The Philosophy of G. E. Moore, Paul Arthur Schilpp (ed.). New York: Tudor Publishing Company.
- Nagel, Thomas (1965). "Physicalism." The Philosophical Review, 74, 3 (July), pp. 339-356.
- Nagel, Thomas (1974/1979). "What Is It Like to be a Bat?" The Philosophical Review 88, 4 (October), pp. 435-450.

- Reprinted in Mortal Questions. Cambridge: Cambridge University Press, pp. 165-179.
- Nagel, Thomas (1986). The View From Nowhere. New York: Oxford University Press.
- Neisser, Ulric (1988). "Five Kinds of Self-Knowledge." Philosophical Psychology 1, pp. 35-59.
- Neisser, Ulric, ed. (1994). The Perceived Self. Cambridge: Cambridge University Press.
- Nietzsche, Friedrich (1954). Thus Spoke Zarathustra. In Walter Kaufman (ed.) (trans.) The Portable Nietzsche. New York: The Viking Press.
- Nichols, Shaun (forthcoming). "The Mind's 'I' and the Theory of Mind's 'I': Introspection and Two Concepts of Self." Forthcoming in Philosophical Topics.
- Nichols, Shaun and Stich, Stephen (forthcominga). "Reading One's Own Mind: Self-Awareness and Developmental Psychology." Forthcoming in Working Through Thought, eds. R. Kukla, R. Manning & R. Stainton. Boulder: Westview Press.
- Nichols, Shaun and Stich, Stephen (forthcoming b). "How to Read Your Own Mind: A Cognitive Theory of Self-Consciousness." Forthcoming in Aspects of Consciousness, eds. Q. Smith and A. Jokic. Oxford University Press.
- Nisbett, Richard E. and Wilson, Timothy D. (1977). "Telling More Than We Can Know: Verbal Reports on Mental Processes." Psychological Review, 84, 3, pp. 231-259.
- Nozick, Robert (1981). Philosophical Explanations. Cambridge: The Belknap Press of Harvard University Press.
- O'Brien, Lucy (1994). "Anscombe and the Self-Reference Rule." Analysis 54, pp. 277-281.
- O'Brien, Lucy (1994-95). "The Problem of Self-Identification." Proceedings of the Aristotelian Society 95, pp. 235-251.

- O'Brien, Lucy (1995). "Evans on Self-Identification." Noûs 29, pp. 232-247.
- O'Brien, Lucy (1996). "Solipsism and Self-Reference." European Journal of Philosophy, 4, 2, pp. 175-194.
- O'Shaughnessy, Brian (1995). "Proprioception and the Body Image." In Bermúdez et al (1995), pp. 175-204.
- O'Shaughnessy, Brian (1980). The Will: A Dual Aspect Theory, 2 vols. Cambridge: Cambridge University Press.
- Parker, Sue T., Mitchell, Robert W., and Boccia, Maria L. (eds.) (1994a). Self-Awareness in Animals and Humans: Developmental Perspectives. Cambridge: Cambridge University Press.
- Parker, Sue T., Mitchell, Robert W., and Boccia, Maria L. (1994b). "Evolving Self-Awareness." In Parker et al, pp. 413-428.
- Pavani, Francesco, Spence, Charles, and Driver, Jon (2000). "Visual Capture of Touch: Out-of-the-Body Experiences with Rubber Gloves." Psychological Science, 11, 5, pp. 353-359.
- Peacocke, Christopher (1983). Sense and Content. Oxford: Oxford University Press.
- Peacocke, Christopher (1992). A Study of Concepts. Cambridge: The MIT Press.
- Peacocke, Christopher (1999). Being Known. Oxford: Clarendon Press.
- Peacocke, Christopher (1998). "Conscious Attitudes, Attention and Self-Knowledge." In McDonald, et al (1998).
- Pears, David F. (1987). The False Prison: A Study of the Development of Wittgenstein's Philosophy, Vols 1-2. Oxford: The Clarendon Press.
- Perry, John (1977) "Frege on Demonstratives." The Philosophical Review, 86 (October), pp. 474-97.
- Perry, John (1979). "The Problem of the Essential Indexical." Noûs 13, pp. 3-21.

- Perry, John (1998). "Indexicals, Contexts and Unarticulated Constituents." Proceedings of the 1995 CSLI-Amsterdam Logic, Language and Computation Conference. Stanford: CSLI Publications.
- Perry, John (2000). The Problem of the Essential Indexical and Other Essays. Stanford: CSLI.
- Platek, Steven M. and Gallup, Gordon G., Jr. (2002). "Self-Face Recognition Is Affected by Schizotypal Personality Traits." Schizophrenia Research, 57, pp. 81-85.
- Povinelli, Daniel J. (1989). "Failure to Find Self-Recognition in Asian Elephants (Elephas Maximus) in Contrast to Their Use of Mirror Cues to Discover Hidden Food." Journal of Comparative Psychology 103, 2, pp. 122-131.
- Povinelli, Daniel J., Rulf, Alyssa B., Landau, Keli R., and Bierschwale, Donna T. (1993). "Self-Recognition in Chimpanzees (Pan troglodytes): Distribution, Ontogeny, and Patterns of Emergence." Journal of Comparative Psychology 107, 4, pp. 347-372.
- Povinelli, Daniel J., and Prince, Christopher G. (1998). "When Self Met Other." In Ferrari and Sternberg (1998), pp. 37-107.
- Proust, J. (1999). "Self Model and Schizophrenia." Cognition 8, pp. 378-384.
- Pryor, James (1999). "Immunity to Error Through Misidentification." Philosophical Topics 26, 1 & 2, pp. 271-304.
- Quine, W. V. O. (1953). From a Logical Point of View. Cambridge: Harvard University Press.
- Quine, W. V. O. (1966). Ways of Paradox. Cambridge: Harvard University Press.
- Quine, W. V. O. (1981). Theories and Things. Cambridge: Belknap Press.
- Reiss, Diana, and Morino, Lori (2001). "Mirror Self-Recognition in the Bottlenose Dolphin: A Case of

- Cognitive Convergence." PNAS 98, 10, pp. pp. 5937-5942.
- Rochat, Philippe, and Striano, Tricia (2002). "Who's in the Mirror? Self-Other Discrimination in Specular Images by Four- and Nine-Month-Old Infants." Child Development, 73, 1, pp. 35-46.
- Rosenthal, David M. (1986a). "Intentionality." Midwest Studies in Philosophy X, pp. 151-184.
- Rosenthal, David M. (1986b). "Two Concepts of Consciousness." Philosophical Studies 49, 3 (May), pp. 329-359.
- Rosenthal, David M. (1993). "Thinking that One Thinks." In Davies and Humphreys eds. Consciousness: Psychological and Philosophical Essays. Oxford: Basil Blackwell, pp. 197-223.
- Rosenthal, David M. (1995). "Moore's Paradox and Consciousness." Philosophical Perspectives 9, pp. 313-333.
- Rosenthal, David M. (2000). "Introspection and Self-Interpretation." Philosophical Topics 28, 2, pp. 201-233.
- Rosenthal, David M. (2002). "Moore's Paradox and Crimmins's Case." Analysis 62, 2, pp. 167-171.
- Rosenthal, David M. (2003). "Unity of Consciousness and the Self." Proceedings of the Aristotelian Society, 103, 3, pp. 325-352.
- Rosenthal, David M. (forthcoming 2004). Consciousness and Mind. Oxford: Clarendon Press.
- Rovane, Carol (1987). "The Epistemology of First-Person Reference." The Journal of Philosophy 84, 3, pp. 147-167.
- Russell, Bertrand (1948). Human Knowledge: Its Scope and Limits. London: George Allen & Unwin.
- Ryle, Gilbert (1949) The Concept of Mind. London: Hutchinson.

- Sacks, Oliver (1984). A Leg to Stand On. New York: Harper and Row.
- Sacks, Oliver (1985). The Man Who Mistook His Wife for a Hat and Other Clinical Tales. New York: Summit Books.
- Sarte, Jean Paul (1968). Being and Nothingness: An Essay on Phenomenological Ontology. Hazel E. Barnes (Trans.). New York: Philosophical Library.
- Searle, John R. (1992). The Rediscovery of the Mind. Cambridge: MIT Press.
- Shelton, Richard (1982). Selected Poems, 1969-1981. Pittsburg: University of Pittsburg Press.
- Shoemaker, Sydney (1963). Self-Knowledge and Self-Identity. Ithaca: Cornell University Press.
- Shoemaker, Sydney (1968). "Self-Reference and Self-Awareness." Journal of Philosophy, vol. 65, no. 19, pp. 555-67. Reprinted in Shoemaker (1996), pp. 3-24.
- Shoemaker, Sydney (1970). "Persons and Their Pasts." American Philosophical Quarterly vol. 7, no. 4, pp. 269-285. Reprinted in Shoemaker (1984), pp. 19-48.
- Shoemaker, Sydney (1984). Identity, Cause, and Mind. Cambridge: Cambridge University Press.
- Shoemaker, Sydney (1986). "Introspection and the Self." Midwest Studies in Philosophy, X, pp. 101-120. Reprinted in Shoemaker (1996), pp. 3-24.
- Shoemaker, Sydney (1994). "Self-Knowledge and 'Inner Sense'." Philosophy and Phenomenological Research 54, pp. 249-314. Reprinted in Shoemaker (1996), pp. 201-268.
- Shoemaker, Sydney (1996). The First-Person Perspective and Other Essays. New York: Cambridge University Press.
- Shumaker, Robert W., and Swartz, Karyl B. (2002). "When Traditional Methodologies Fail: Cognitive Studies of Great Apes." In Bekoff et al (2002), pp. 335-343.
- Snodgrass, Joan G. and Thompson, Robert L. (eds.) (1997). The Self Across Psychology: Self-Recognition, Self-

- Awareness, and the Self Concept. New York: New York Academy of Sciences.
- Stalnaker, Robert C. (1991). "Indexical Belief." Synthese, 49, 1 (October), pp. 129-151.
- Stephens, G. Lynn and Graham, George (1994). "Self-Consciousness, Mental Agency, and the Clinical Psychopathology of Thought Insertion." Philosophy, Psychiatry, and Psychology 1, pp. 1-10.
- Stephens, G. Lynn and Graham, George (2000). When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts. Cambridge: MIT Press.
- Stone, T. and Young, A. W. (1997). "Delusions and Brain Injury: The Philosophy and Psychology of Belief." Mind and Language 12, pp. 327-364.
- Strawson, Galen (1999). "The Self and the SESMET." Journal of Consciousness Studies, 6, 4 (April), pp. 99-135.
- Strawson, P. F. (1959). Individuals. New York: Routledge.
- Strawson, P. F. (1994). "The First Person--and Others." In Cassam (1994), pp. 210-215.
- Suarez, S. and Gallup, Gordon (1991). "Self-Recognition in Chimpanzees and Orangutans, but not Gorillas." Journal of Human Evolution, 10, pp. 175-188.
- Swartz, Karyl B. (1997). "What Is Mirror Self-Recognition in Nonhuman Primates, and What Is It Not?" In Snodgrass and Thompson (eds.), pp. 65-71.
- Swartz, Karyl B. (1998). "Self-recognition in Non-human Primates." In Parker et al (eds.), pp. 189-206.
- Swartz, Karyl B., and Evans, Siân (1991). "Not all Chimpanzees (Pan Troglodytes) Show Self-Recognition." Primates 32, pp. 483-496.
- Swartz, Karyl B. and Evans, Siân (1994). "Social and Cognitive Factors in Chimpanzee and Gorilla Mirror Behavior and Self-Recognition." In Parker et al, pp. 189-206.

- Swartz, Karyl B., Sarauw, Dena, and Evans, Siân (1999). "Comparative Aspects of Mirror Self-Recognition in Great Apes." In The Mentalities of Gorillas and Orangutans in Comparative Perspective, S. T. Parker, R. W. Mitchell (eds.), and H. L. Miles, pp. 283-294. Cambridge, England: Cambridge University Press.
- Taschek, William (1985). "Referring to Oneself." Canadian Journal of Philosophy, 15, 4, pp. 629-652.
- Taylor, C. (1985). "Self-Interpreting Animals." In Human Agency and Language: Philosophical Papers, Vol. 1. Cambridge: Cambridge University Press.
- Thompson, Robert L. and Boatright-Horowitz, Susan L. (1994). "The Question of Mirror-Mediated Self-Recognition in Apes and Monkeys: Some New Results and Reservations." In Parker et al, pp. 350-360.
- Tomkins, Calvin (1996). Duchamp: A Biography. New York: Henry Holt and Company.
- van den Box, Esther, and Jeannerod, Marc (2002). "Sense of Body and Sense of Action Both Contribute to Self-Recognition." Cognition 85, pp. 177-187.
- Van Straaten, Zak, ed. (1980). Philosophical Subjects: Essays Presented to P. F. Strawson. Oxford: Oxford University Press/Clarendon Press.
- Vendler, Zeno (1972). Res Cogitans. Ithaca: Cornell University Press.
- Wann, J. P. and Ibrahim, S. F. (1992). "Does Limb Proprioception Drift?" Experimental Brain Research 91, pp. 162-166.
- Wegner, Daniel M. (2002). The Illusion of Conscious Will. Cambridge: Bradford Books/MIT Press.
- Wegner, Daniel M. and Wheatley, T. (1999). "Apparent Mental Causation: Sources of the Experience of Will." American Psychologist, 54, pp. 480-491.
- Weiskrantz, Lawrence (1986). Blindsight. A Case Study and Implications. Oxford: Oxford University Press.

- Weiskrantz, Lawrence (1997). Consciousness Lost and Found: A Neuropsychological Exploration. Oxford: Oxford University Press.
- White, Peter A. (1988). "Knowing More About What We Can Tell: 'Introspective Access' and Causal Report Accuracy 10 Years Later." British Journal of Psychology, 79, pp. 13-45.
- Wilson, Timothy D., Sara D. Hodges, and Suzanne J. LaFleur (1995). "Effects of Introspecting About Reasons: Inferring Attitudes from Accessible Thoughts." Journal of Personality and Social Psychology, 69, pp. 16-28.
- Wittgenstein, Ludwig (1961). Tractatus Logico-Philosophicus. Trans. D. F. Pears and B. F. McGuinness. London: Routledge & Kegan Paul.
- Wittgenstein, Ludwig (1968/1993). "Notes for Lectures of 'Private Experience' and 'Sense Data'," edited, with some translation, by Rush Rhees. The Philosophical Review, 77, pp. 275-320; reprinted in Philosophical Occasions, 1912-1951. Indianapolis: Hackett Publishing Company, Inc.
- Wittgenstein, Ludwig (1969). The Blue and Brown Books. Oxford: Oxford University Press/Basil Blackwell.
- Wittgenstein, Ludwig (1970). Philosophical Investigations. Trans. G. E. M. Anscombe, third ed. New York: The Macmillan Company.
- Young, Andrew W. (1998). Face and Mind. Oxford: Clarendon Press.
- Young, Andrew W. (1999). "Delusions." The Monist, vol. 82, no. 4, pp. 571-589.
- Young, Andrew W. (2000). "Wondrous Strange." Mind and Language, vol. 15, no. 1, pp.
- Young, A. W., Hay, D. C., and Ellis, A. W. (1985). "The Faces that Launched a Thousand Slips: Everyday Difficulties and Errors in Recognizing People." British Journal of Psychology, 76, pp. 495-523. Reprinted in Young (1998).

Zahavi, Dan (2000a). "Self and Consciousness." In D. Zahavi (2000) , pp. 55-74.

Zahavi, Dan (2000). Exploring the Self. Amsterdam and Philadelphia: John Benjamins Publishing Company.