

SELVES AND OTHERS:
AN INTERPERSONAL ACCOUNT OF SELF-CONSCIOUSNESS

by

JAMES M. DOW

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

2011

©2011

JAMES M. DOW

All Rights Reserved

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

Barbara Montero

Date

Chair of Examining Committee

Iakovos Vasiliou

Date

Executive Officer

Advisor: Jesse Prinz

Supervisory Committee:

Michael Devitt

Barbara Montero

David Rosenthal

Axel Seemann

THE CITY UNIVERSITY OF NEW YORK

Abstract

Selves and Others: An Interpersonal Account of Self-Consciousness

by

James M. Dow

Advisor: Professor Jesse J. Prinz

My dissertation presents an argument for the claim that awareness of oneself and awareness of others is symmetrical and mutually dependent. My work challenges the traditional account of self-consciousness according to which individuals can be aware of themselves even though they have never been aware of individuals like themselves.

First, I provide an analysis of self-consciousness as the self-ascription of experiences that shows that if a subject is to be able to think “I am experiencing F,” then he must be able to ascribe experiential predicates, e.g., “b is F,” “c is F,” to arbitrarily distinguishable individuals.

Second, I argue that in order for one to be self-conscious, one must be able to identify oneself as a subject of experience. However, the traditional account of self-ascription holds that self-ascriptions do not involve identification of a subject, because ‘I’ is immune to error through misidentification. Contrary to universal opinion, I argue that self-ascriptions are not immune to error through misidentification through a conceptual and empirical argument.

Third, I argue that the identification of the subject of self-ascription is only possible given the perception of oneself as a person among persons, which I call *the Persons Theory*. The Persons Theory provides us with a genuinely unique account of thought about other minds that differs from two extant accounts of experience

ascription— the simulation theory and the theory-theory. According to the Persons Theory, rather than imagination or thought, *perception* of persons enables the self-ascription and other-ascription of experiences. I elucidate types of recognition and acknowledgement between subjects in joint perception, action and emotion that are pivotal for self-awareness. An implication of the Persons Theory is that awareness of oneself and awareness of others develops in tandem and involves interaction between persons.

Acknowledgements

This dissertation is definitely the product of much interpersonal interaction. My ideas presented here are the result of many conversations, lectures, presentations, and passing remarks. Some of the ideas presented here were presented formally to audiences, but much of the formation of this project has arisen through informal and casual conversation. For this reason, it is difficult to pay debt and express gratitude to everyone that has helped along the way. However, I will do my best to thank those that have obviously contributed to the development of this project.

Jesse Prinz has been an incredible advisor, professional mentor, and a steadfast source of inspiration. I could not have made it through this process without his encouragement, sage advice, and constructive criticism. He is an exemplar of what philosophical advising should be— a balance between being carefully receptive and respectfully critical. I am deeply grateful for his help in advising this project.

I would like to thank David Rosenthal for inspiring me to pursue a project in philosophy of mind and consciousness in general and for encouraging me to focus more deeply on the self-other relation. It was his claim that I could make a career out of writing about self-consciousness and intersubjectivity that inspired me to take up the project seriously. I am grateful to Barbara Montero for reading and commenting on the dissertation and for inviting me to audit a class on bodily self-awareness. Several discussions with Barbara also enabled me to put my work in more commonsense, everyday language.

I want to thank Axel Seemann for inviting me to Bentley University for a conference on joint engagement, at which I presented material from the dissertation.

Axel also gave extensive comments on a draft of the dissertation and other writing, enabling me to improve my arguments and claims and pointing my research in new and challenging directions. I am grateful to have found in Axel an ally in the study of person perception and social cognition research.

I want to express profound gratitude to my best friend, Gabriel Gottlieb. He has read every page of this dissertation. His summaries, critiques and comments definitely improved the final product. The project could not have been completed without his enduring encouragement, Saturday morning café discussions, and late night “red light” phone conversations. I am thankful to have a companion in the pursuit of an interpersonal account of self-consciousness that respects historical concerns, analysis and interdisciplinary work in the sciences. I am also grateful to have such a dedicated friend.

I would also like to thank my friends at the CUNY Graduate Center that helped in numerous ways. Carl Hammer, Myrto Mylopoulus, Dan Shargel, Peter Langland-Hassan, Ellen Fridland, David Pereplyotchik, and Chris Allen Sula all helped in the development of my ideas. I would also like to thank my friends at Columbia University, Michael Brent and Avery Archer, whose conversations were pivotal in enabling me to understand my project. And, I would like to thank my friends from the New School for Social Research, Adam Gies and Mark Theunissen for their helpful commentaries on my work on immunity to error through misidentification.

I dedicate this dissertation to my parents, Debra Piccirillo and Gary Dow. They have given me moral and financial support throughout my undergraduate and graduate years, without which I would not have made it through. I am grateful to my mother for always encouraging me to ‘be myself,’ for keeping me steadfastly on the path, and for

being there to share my experiences. I am grateful to my father for being there to listen to my difficulties, to offer inspiring words, and for providing me with a summer retreat in the Black Hills of South Dakota, where most of this project was written. I would not be who, what or where I am without my parents, and I am thankful for the upbringing that I received.

I would also like to thank Steve Piccirillo, who has been a second father to me, providing me with steadfast reassurance and support through the ups and downs throughout my education (Go Yankees!). I also want to thank my Mother-in-Law, Melody Cowper-Smith, who at a time that I was feeling rudderless in my project, steered me in the right direction because of her careful ability to listen.

Last but not least, I want to express my love and gratitude to my wife, Melissa Cowper-Smith. She has been a source of support and inspiration throughout this entire process. On countless occasions, she talked me through the difficulties, never letting me quit. I have profound respect for her strength, her passion and her authenticity. Throughout the years, she provided an environment that enabled me to get work done. But, more importantly, when I needed to escape, she was always there to go camping, hiking and walking together in the woods. Those adventures with her got me through this process. On the day I defended this project, she was nine months pregnant and having a procedure done in the hospital. Despite being afraid and overwhelmed, she was still a source of support and encouragement. Two weeks later, our beautiful son, Afton was born. I am grateful to be married to such a strong, beautiful and natural woman, and I look forward to the joint experience of bringing up our wonderful boy.

Table of Contents

INTRODUCTION

The Solo Question

Against Cartesian Individualism about Self-Consciousness	1
1. The Pernicious Influence of Cartesian Individualism	4
2. Historical Arguments for the Symmetry Thesis	20
3. The Constructive Argument for the Symmetry Thesis	28

CHAPTER ONE

The I is a We; the We, an I

Constitutive Claims, Development and Transcendental Arguments	32
1. Bermúdez’s Constitution Claim	43
2. How-Possible Questions about Self-Ascription	60
3. Transcendental Arguments	68

CHAPTER TWO

I See a Sheep

Self-Ascription, the Subject-Predicate Structure and Generality	80
1. Models of Self-Ascription, Hume’s Error and Neo-Expressivism	82
2. Self-Ascription and Subject-Predicate Structure	102
3. The Subject-Predicate Structure and the Generality Constraint	119

CHAPTER THREE

Shoegenstein on Self-Ascription, Immunity to Error and the I-as-Subject	132
1. Wittgenstein on the Possibility of Error and I-as-Subject Uses	135
2. Shoemaker on IEM and Self-Presentation “From the Inside”	137

3. Shoegenstein is not Wittgenstein	143
4. No Non-Circular Account of the SC _S /SC _O Distinction	148
5. Objections	161
6. Self-Ascription without the Dogmas	163

CHAPTER FOUR

Subject to Error

Conceptual Arguments and Empirical Evidence to Deny IEM	168
1. Self-ascriptions are not IEM	168
2. The Conceptual Argument Against IEM	171
3. The Empirical Argument Against IEM	176
4. Self-Ascription Without the Dogmas	190

CHAPTER FIVE

Against Cognitive Descriptivism

Self-Ascription, Identification and the Persons Principle	194
1. The Target: The Lockean View	197
2. Egos, Functional Unity and Lived Embodiment	200
3. For the Persons Principle: A Kantian Argument	211

CHAPTER SIX

On the Joint Engagement of Persons

Self-Consciousness, the Symmetry Thesis and Person Perception	217
1. Introduction	217
2. The Neo-Lockean Account of Joint Engagement	222
3. The Persons Theory of Joint Engagement	232

4. Objections to the Persons Theory	249
5. Concluding Remarks	251
CHAPTER SEVEN	
<i>Dissolving the Conceptual Problem of Other Minds</i>	
Self-Ascription, Symmetry and Person Perception	260
BIBLIOGRAPHY	276

Introduction

The Solo Question: Against Cartesian Individualism about Self-Consciousness

Many philosophical accounts of the nature of self-consciousness are committed to the idea that self-consciousness and the self should be understood as fundamentally individualistic. An individualistic account of self-consciousness represents an individual as consciously aware of itself— whether that involves experiencing, thinking, or perceiving itself— even though that individual might exist in solitude and isolation. Occasionally, philosophers of mind present explicit arguments for individualism; however, for the most part, individualism remains merely a default theoretical position. The view that self-consciousness does not require awareness of other self-conscious beings is often supported by imaginative stories— Robinson Crusoe, Kaspar Hauser, and other stories about solitary individuals— which depict a self-conscious individual living on a desert island or otherwise in solitude and isolation, never interacting with other creatures. The assumption that is implicit in these narratives is that self-consciousness is innate, inborn, or a basic feature of the mind, one that can occur even in the absence of others.

Let's call the fictional hero of these stories 'Solo' (A name coined by Galen Strawson (2008, p. 122).) If we think that Solo would be capable of self-consciousness, then he might be described as thinking something like the following, "I am only conscious of myself. That which I am conscious of is my experience and my experience is the experience of no other. I am the only self in existence and my consciousness is the only consciousness in existence." (Galen Strawson (2008) introduces the idea of 'Solo' to defend an individualistic notion of self-consciousness.) Let's consider a version of the

Solo story to probe the assumptions of individualism about self-consciousness. This will serve as a narrative setup for the dissertation.

After a recent economic depression forced the closure of a biogenetics lab raising human infants from test tubes, a group of scientists were forced to abandon Solo as a one-week-old infant. He was sent to a wilderness area with the idea that he would be consumed by mountain lions. Fortunately, assisted by a variety of accidental events and abnormal fortitude, Solo was able to survive. He learned to find shelter, forage and hunt for food, collect water, and surprisingly escaped the scythe of the reaper several times. A few questions arise: Assuming that Solo has never interacted with other self-conscious individuals like himself, is he capable of self-consciousness? Is such an isolated child capable of entertaining consciousness of himself as himself?

The answer to this question partly depends on what we think about the nature of self-consciousness. Opinions concerning the nature of self-consciousness vary widely. However, most philosophers and psychologists mean by ‘self-consciousness’ something like the ascription of experiences to oneself as a subject. According to this account, one is self-conscious if for instance, one is thinking of oneself that one is seeing a wolf-child, or that one is dancing with wolves, or that one is afraid of the wildfire burning on the hill. In these cases, one is ascribing an experience— a perception, an action, and an emotion— to oneself. Although this is certainly not uncontroversial, I will take this as a starting point for discussion about self-consciousness, because the figures that are central for the discussion (P. F. Strawson (1959; 1966a), Gareth Evans (1982), Quassim Cassam (1997), and José Luis Bermúdez (1998)) all suppose that self-ascription is a central

notion of self-consciousness. The notion of self-consciousness as self-ascription will take center stage in Chapter 2.

The self-ascription of experiences is familiar to normal adult human beings. For instance, we are familiar with saying and thinking thoughts like the following: “I see a wolf-child” (a self-ascription of perception), “I am dancing with my wolf-child” (a self-ascription of action), or “I am envious of my wolf-child’s primordial comportment” (a self-ascription of emotion). Given the focus is self-ascription of experiences, then the question raised above can be rephrased in the following way: *Assuming that Solo has never interacted with other individuals like himself, is he capable of self-ascribing experiences?* (Call this ‘the Solo question.’)

What is puzzling is that there is a divergence in answers between disciplines of philosophy and psychology. Many philosophers of mind (mostly implicitly) respond that a self-conscious Solo is possible (or at least not conceptually impossible). A short perusal of recent anthologies of self-consciousness suffices to show that the self-other relation receives short shrift in accounts of self-ascription (Cassam (1994); Eilan and Roessler (2003); Bermúdez, Marcel, and Eilan (2003); Brook and Devidi (2001)). One counterexample is Thompson (2001). The psychological community, however, especially developmental psychologists, tend to respond explicitly that a self-conscious Solo is impossible (Trevarthen (1979); Tomasello (1993); Rochat (2009)).

Many developmental psychologists take it as a basic assumption that self-consciousness as self-ascription is dependent developmentally up on non-negligible interpersonal interaction. To be aware of oneself requires being aware of others from the earliest moments of infancy, and even while in the womb. The interpersonal interaction

that I have in mind is jointly perceiving an object, jointly intending to do something or other, and jointly responding emotionally to objects in one's environment. (Some anthologies that approach the issue of joint engagement include Moore and Dunham (1995); Eilan, Hoerl, McCormack, and Roessler (2005); and Seemann (2010). This literature will take center stage in Chapter 6.)

However, notwithstanding this divergence in opinion, the philosophical and the psychological camps agree that we can conceive, imagine or entertain the idea that such a creature as Solo exists. Our capacity to conceive and imagine such narrative possibilities may lead us to think that being self-conscious (and being a self) is available to a creature although that creature is entirely alone in the world. We might believe we can conceive or imagine such thing because we can imagine— inspired by the Twilight Zone— a great subtraction of individuals from the earth, whence all that remains is one individual. One might think that if this is possible to imagine, then it is equally possible to imagine any one creature being a self-conscious being even if there never were, are or will be any other self-conscious beings in existence. However, my goal in the dissertation is to challenge the import of this type of conceivability or imaginability argument. Despite our fondness of such fictional possibilities, I will argue in the dissertation that we should reject individualism and answer the Solo question with an unequivocally negative response. However, I will not approach the conceivability or imaginability argument directly, because I do not think anything profitable arises from approaching such arguments directly. Instead, my position tends towards an alliance with the developmental psychologist rather than with the contemporary trends in philosophy.

1. The Pernicious Influence of Cartesian Individualism

However, even if one rejects the conceivability arguments, one still has to approach the question of why philosophers are so gripped by individualism about self-consciousness. Why are we so ready to conceive or imagine that a self-conscious Solo is possible? Apparently, Descartes would have held the idea that Solo was possible given the individualism that is implicit in the Discourse on Method and The Meditations. The I think, or ego, or self that is revealed in the reflection of “Cogito, Ergo Sum” is a solitary consciousness, a thinking substance that thinks about itself and its conscious states. According to the Cartesian picture, we think only problematically about our own bodies, the external world, or other subjects of experience. I say ‘problematically,’ because according to Descartes, thinking about our own bodies requires solving the mind-body problem, thinking about the external world requires determining that a non-deceiving God exists. What is required in order to overcome the problem of other minds?

Central to our purposes, thinking about others requires overcoming a genuine metaphysical divide between our own minds (experienced directly and immediately) and the minds of others (experienced indirectly and mediately, if at all). Descartes might argue that the individualistic account of self-consciousness is defensible, because we can conceive and imagine that Solo could exist even if no other subjects of experience existed. There has been much argument trying to overturn Cartesian philosophy of mind in the first two respects. For instance, philosophers have argued against Cartesian mind-body dualism (Good recent anthologies are Rosenthal (2000) and Crane (2001)). Some have argued for the idea that the mind and mental states depends for its significance upon material objects in space and time (Kant (1781); Strawson (1959); Evans (1982); Campbell (1994)). Others have argued that selves are in important respects materially

embodied (Cassam (1997); Gallagher and Shear (1999)). At least since Kant, the Cartesian problem of external world has received a variety of solutions. All of which is to say that there has been consistent focus on the mind-body problem and the problem of the external world in the late twentieth century.

However, there has been little systematic critique of Descartes' individualist account of self-consciousness. This may be partly because the trend in philosophy of mind in general has been to focus on consciousness. In the last few decades, we have seen a lot of ink spilled about consciousness; the topic of self-consciousness is only recently getting direct attention in philosophy (One exception is Sebastian Rodl's (2007) book Self-Consciousness). However, even in the books and articles with specific focus on self-consciousness, the question of whether self-consciousness is related to consciousness of other beings, creatures, or subjects has taken a back seat to other concerns. The focus of the literature about self-consciousness has been variously concerned with such topics as: non-circular and non-regressive accounts of self-reference (Henrich (1971), Nozick (1981)); the relation between self-consciousness and immunity to error through misidentification (Kriegel (2008); Smith (2006)); the role that self-consciousness plays in perception and action (Hurley (1998), Noe (2004; 2009); O'Brien (2007)), or the feasibility of arguments for materialism about self-consciousness (Cassam (1994), Smith (Diss)). While these topics play a role in the dissertation, there is not much critique of the Cartesian individualist position in this literature. In this sense, I have a modest hope of inspiring a new wave of focus on anti-individualism about self-consciousness. I do not plan to focus in this dissertation on Descartes' philosophy of

self-consciousness (instead, on what a Cartesian individualist position might be), but I do take my project to be fundamentally anti-Cartesian and anti-individualistic.

Now that I have positioned my project with respect to the history and the recent literature on self-consciousness, I will turn to expressing the goal of the project. My goal in the dissertation is present a constructive argument for anti-individualism about self-consciousness as self-ascription that draws on the work of P. F. Strawson, Gareth Evans and J. L. Bermúdez. Each of these figures' work can be understood as overcoming Cartesian individualism about self-ascription. My goal is to present an argument that defends what I will call 'the Symmetry Thesis,' (following Bermúdez (1998)) which proposes that in order for an individual to be capable of self-ascribing experiences, that individual must be capable of ascribing experiences to other creatures like itself (cf. Bermúdez (1998) for differences between our statements of *the Symmetry Thesis*). How does this thesis militate against Cartesian individualism about self-consciousness? In order to make this clear, I now discuss what I take Cartesian individualism about self-consciousness to be.

The Cartesian individualistic account of self-consciousness maintains that the Cogito's nature enables any creature possessing a mind to entertain a relation to oneself as oneself. The innate human mind itself determines that one is capable of thinking "I am, I exist" (p. 64). According to Descartes, Solo would be able to entertain this relation to himself as himself, despite having never interacted with— whether that involves sensing, perceiving, thinking about, or imagining— other self-conscious beings like himself. Descartes also gives different accounts of the contents that enter into the self-ascriptions of experiences, which affects differences between those self-ascriptions. As

will be made clear in Chapter 2, I think there can be a systematic account of self-ascription presented that abstracts away from these content-specific effects. The Cartesian individualist view of self-consciousness has left an indelible mark on the philosophy of mind. In particular, there are still remnants of Cartesian individualism concerning self-consciousness.

For instance, Galen Strawson (2008) has argued that Solo, despite having never interacted with an other being like itself, would be capable self-consciousness: “There is no obvious [or unobvious] incoherence in the idea of a subject of experience S that thinks of itself in the distinctively self-conscious way, and uses a first person pronoun like ‘I,’ but has... no belief in other subjects of experience” (2008, p. 122). According to G. Strawson, there are no good arguments nor any evidence for the idea that in order for an individual to be self-conscious, that individual needs to be aware of creatures like itself. Solo need not believe there are others, conceive of others, or otherwise interact with others in any form. Notwithstanding this complete lack of others, Solo would possess self-consciousness. In this dissertation, I will challenge this Cartesian individualist view. (Philosophers such as Thomas Nagel (1974; 1986) and John Searle (1983; 2004) are also each committed to Cartesian individualism about self-consciousness.) My argument here can be understood in terms of an argument between “Strawson the Younger” (Galen Strawson), who argues that the Symmetry Thesis (what he calls ‘The Others Thesis’) is false, and “Strawson the Elder” (Peter F. Strawson), who argues that the Symmetry Thesis is true. I am interested in defending the position that Strawson the Elder’s position should be preferred.

The constructive argument for *the Symmetry Thesis* concludes that self-consciousness as self-ascription is dependent upon awareness of others. So, on the assumption that Solo was not aware of others, he could not be aware of himself as himself. Before I begin to make this argument, I need to make clear and precise what the opposition believes, which requires a discussion of the commitments of an individualistic account of self-consciousness.

One way to make clear what individualism about self-consciousness involves is to focus on its connection to the classic position called ‘solipsism’ (from the Latin *solus ipse*— oneself alone). Solipsism is a position about the existence of a subject of consciousness; solipsism maintains that the only being in existence is oneself. There are two different ways to describe the position of solipsism. The first description suggests that an individual is the only self, or locus of consciousness in existence. This position denies the existence of other selves or subjects of consciousness. For example, Solo might be surrounded by a world of rocks, plants, animals, and even embodied human beings, but none of these denizens of Solo’s world can be understood as in any way minded individuals. The second description suggests that nothing at all exists apart from an individual’s mind and mental states, which denies the existence of anything apart from oneself and the contents of one’s mind. For example, Solo is not surrounded by anything that is intelligible as an independently existing world at all, but instead, his world is strictly speaking, limited by himself and the contents of his consciousness (Wittgenstein, 1961, §5.62)). Often, both of these descriptions of solipsism are used as curb against extremely bad theories of self-consciousness. It might be argued that if one’s account of self-consciousness entails solipsism, then by *reductio ad absurdum*, one or other tenet of

one's position must be rejected. It is not this type of solipsism that I have in mind when I say that an account of self-consciousness is individualistic, since these metaphysical views of self-consciousness are not my focus. We might call solipsism a metaphysical account of self-consciousness.

Another account of solipsism, however, is called 'methodological solipsism' which suggests that the content of an individual's mental states is completely determined by facts about that individual, specifically, determined by the causal-functional roles that such mental states play in the individual's mind/brain. (Fodor's (1980) account, Stich's (1980) response, and Fodor's Reply (1980), Noonan (1980), Rowlands (1995) are good sources). Methodological solipsism is committed to the idea that the contents of mind/brain states are determined by facts independent of the environment, including the social environment and other individuals. It is this type of individualism about self-consciousness that is my concern. Under this construal, an account of self-consciousness is individualistic if it denies the significance of the social environment— which includes other beings like oneself, or one's conspecifics— for the constitution of the content of self-consciousness. According to methodological solipsism, any proper cognitive psychology of self-consciousness theorizes about the individual independent of environmental and social constitution. However, since the notion of 'methodological solipsism' is tied to a specific research strategy in cognitive psychology, it would take us too far afield to discuss the debates concerning the relative success of that research strategy. I will not be arguing that my constructive argument bears on questions of the viability of that research strategy in general. Instead, my conclusion is more conservative. My claim is that if *the Symmetry Thesis* is true, then Cartesian

individualism qua methodological solipsism *about self-consciousness* is false. What that would mean is that inquiry into self-consciousness should also consider the second person and the third person, because self-consciousness cannot be understood as a solely first person phenomenon.

Individualism in the philosophy of mind is a more general methodology or set of theoretical commitments than methodological solipsism about cognitive psychology. It involves a general attitude towards how we should talk and think about self-consciousness. A number of philosophers have been inspired by Cartesian individualism and support the intuitions that are culled from Solo-style thought experiments. There are also explicit arguments for individualism about self-consciousness, whether or not the Cartesian individualist tradition or the Solo narrative is the chief influence. What is the implicit argument with the Solo narrative and how is it an argument for individualism about self-consciousness? I will analyze what I think is implicit in the discussion of the Solo narrative that is marshaled in favor of individualism about self-consciousness in order to allow the reader to better understand the opposition.

The argument begins with Solo in the wild. The question that is usually asked is whether the relation between Solo and others is necessary. The idea is that if the relation between self and other were necessary, then when Solo introspects about his self-consciousness, then Solo would either be presented with an intentional other, namely a generalized idea of an Other or with the relation between self and other. However, when one introspects about one's self-consciousness, one is presented with oneself, not an other. The suggestion is that if Solo were impossible, then it would be part of the introspectible content of self-consciousness that one is aware of an ominous Thou that is

ever present in one's awareness of oneself. However, according to the individualist, no such experience is available to us, since we are presented only with ourselves in self-consciousness. So, the individualist argues, the left side of this disjunction must be false.

Then, the right side of the disjunction is considered, namely that if one's introspects about one's self-consciousness, then the relation between self and other would be present. Often, there is a distinction made between an intrinsic relation and an extrinsic relation. The argument for individualism describes presentation of an other as *intrinsic* to self-consciousness, if in being aware of oneself, one is (by definition) aware of other subjects. It would follow from the very concept of self that it was related to the other. The argument for individualism describes the presentation of an other as *extrinsic* to self-consciousness, if in being aware of oneself, one would have to *at some point or other* be related to an other subject, although it need not enter into the introspectible content of self-consciousness. While it does not appear introspectively at each instance that one is aware of other subjects, it might be that the relation is nevertheless present to one.

The argument for individualism continues, beginning with the intrinsic relation, introspective self-consciousness does not present an ominous Thou, because when we reflect introspectively, we are aware of ourselves in a way that does not require thinking about, imagining, or perceiving other subjects, so self-consciousness must be individualistic. Continuing with the extrinsic relation, then the argument for individualism suggests that introspective consciousness does not discover an extrinsic relation to other subjects, so self-consciousness must be individualistic. The conclusion of the argument for individualism about self-consciousness, is then because introspective

self-consciousness does not present reasons or evidence for an intrinsic or extrinsic relation between self and other, then therefore, the concept of self-consciousness is individualistic concept. In accounting for the nature of self-consciousness, we need not consider any non-individualistic relations, whether intrinsic or extrinsic.

Before I elucidate the constructive argument for *the Symmetry Thesis* as a response to Cartesian individualism, I want to provide a few objections to the argument for individualism about self-consciousness. The first objection is that introspection of self-consciousness is not sufficient to provide an understanding of the nature of self-consciousness. The second objection is that individualism about self-consciousness is based on assumptions about self-knowledge. The third objection is that the Solo narrative and individualism about self-consciousness does not consider the import of transcendental arguments for the conditions of self-consciousness, but instead focuses on intuitions about self-consciousness.

Objection to Introspection as the Sole Arbiter

My first objection to the argument from individualism is that it privileges introspective self-consciousness. As Galen Strawson writes, “it isn’t just obvious that self-consciousness requires actual possession of some conception of other subjects of experience” (2008, p. 121). However, introspection cannot be supposed to be the sole arbiter (Schwitzgebel (2008)). Self-consciousness is surely a first-personal phenomenon, but that does not mean that other second-personal and third-personal concerns should not be considered.

Another account, which presumes introspection to be the sole arbiter, is Nagel’s account of self-consciousness in The View from Nowhere. He focuses, as I do, on self-

ascription of experiences as central to self-consciousness, when he writes “being mine is an irreducible, unanalyzable characteristic of all my mental states, and that it has no essential connection with anything in the objective order or any connection among those states over time” (1986, p. 34). In Nagel’s arguments about the contrast between subjectivity and objectivity, he focuses on the introspection of experiences in order to argue that there is no intrinsic or extrinsic connection between the first person and the third person. However, throughout his work, he hints at the possibility that the relation between the first person and the second person is relevant (1974). As will become clear in chapter 6, this second-person approach to the study of consciousness and self-consciousness is what has been missing as a response to Nagel’s worries about subjectivity and objectivity.

I do not mean to suggest that we must proceed with a wholly non-introspective account of self-consciousness under which an individual’s perspective on her own phenomenology would not be at all relevant. That would mean that we should approach self-consciousness from an entirely third-personal perspective. While I think this approach is merited in the long run, there is still work to be done in articulating our commonsense understanding of self-consciousness. As will become clear in chapter 2, I favor a functionalist account of self-consciousness understood in terms of self-ascription and self-ascription involves the positing of I-thoughts. My argument is that the positing of I-thoughts is done not only from the first person perspective, but also from what I will call ‘the second-person perspective.’

Objection to Self-Knowledge as the Focus

My second objection to the argument for individualism is the claim that it presupposes certain claims about self-knowledge that have been questioned in philosophical debates about other minds. For instance, since Thomas Reid's insightful formulation of the problem of other minds, philosophers have addressed the issue of Cartesian individualism by arguing that we possess knowledge of other minds or that our concept of other minds is legitimate. However, in its epistemological and conceptual forms, the problem has focused predominantly on the contents of minds in general (Avramides (2001)). Rather than concentrating specifically on self-consciousness as the central phenomenon, the problem of other minds has focused on knowledge of the contents of mental states and conditions of the ascription of those contents, for instance, the ascription of pains or headaches. The problem mostly focuses on our knowledge of other minds, rather than methodological approaches to the contents of mental states or processes, in particular on the content of self-consciousness.

For instance, although Fodor is willing to accept that his philosophical upbringing involved much haranguing about the problem of other minds, he suggests, "It's gotten hard to believe that there is a special problem about the knowledge of other minds (as opposed to knowledge of anything else)" (1994, p. 294). However, despite Fodor's implication that there is not a legitimate problem about the knowledge of other minds, the problem of other minds is conceptual in nature, not epistemological. I will argue in the dissertation that if *the Symmetry Thesis* is correct, then our inquiry into self-consciousness should not allow a theoretical rent between self and other. I claim that if this theoretical rent is not allowed to open, then the problem never arises as a problem concerning the differences between self-knowledge and other-knowledge. In Chapter 7, I

will discuss how the argument for *the Symmetry Thesis* enables us to overcome the conceptual problem of other minds.

Avramides (2001) has suggested that Fodor's weak dismissal of the problem of other minds leaves him committed to Cartesian Individualism. Instead of accepting Cartesian individualism, we should "reject the Cartesian framework; we must understand the conceptual assumptions that shape and give rise to our epistemological questions" (2001, p. 16). I agree with Avramides that the proper trajectory would involve attacking Cartesian Individualism at its heart by treating the conceptual problem of other minds to be generated by basic assumptions about the nature of self-consciousness.

Avramides also notes that the problem of other minds has found a slight resurgence in the recent inquiry into self-knowledge and externalism. (Ludlow and Martin (1998) and Nuccitelli (2003) are two anthologies that are particularly instructive.) However, these debates are not directly focused on self-consciousness as such; instead, they are focused on what the status of our knowledge of the content of our mental states is, supposing that the content of those mental states are external. Mostly, the concern is whether self-knowledge is compatible with externalism about mental contents. In this sense, there have been challenges to Cartesian individualism. However, as I mentioned above, I would reject the assumption that introspection is the sole arbiter, and this assumption is usually used to generate the problem about self-knowledge and externalism.

The arguments against individualism, or what have been called 'anti-individualist arguments' in Putnam (1975) and Burge (1979) focus more specifically, on the influence that the natural and social environment has on mental content in general, rather than

focusing on the particular mental content of self-consciousness. But, even Putnam and Burge presuppose individualism about self-consciousness, since both assume that self-consciousness should admit narrow contents. I should make clear, then that the concern of the dissertation is not with self-knowledge and the question of its incompatibility with externalism about mental content, but instead with the nature of self-consciousness. Neither is the concern with the arguments for content externalism taken up by Burge and Putnam (Although, I do consider this literature in a paper titled ‘Self-Consciousness Ain’t in the Head: An Argument for the Vehicular Externalism of Self-Ascription’.)

The Objection to the Use of Intuitions

My third objection to the argument for individualism about self-consciousness brings us directly to the constructive argument for *the Symmetry Thesis*. The third objection says that we should not merely play tennis with intuitions about the nature of self-consciousness, but instead should consider arguments and evidence for the conditions of self-consciousness. After all, it might be suggested against the individualist’s intuitions that it is equally conceivable that our desert island infant Solo lacks self-consciousness. It is further conceivable that Solo lacks other mental states, such as consciousness, linguistic capacities, the capacity to judge, believe or think, or even such basic or primitive states as moods, emotions, or feelings. We can equally conceive that this infant with a human body and brain crawling about among the coconuts like a snake or a lizard. We can describe this primordial creature as entertaining only those states and processes necessary for rummaging for edible insects and plants.

So, while it may be the case that Cartesian individualists possess the intuitions that a creature can be self-conscious notwithstanding its utter isolation, it may also be the

case that anti-Cartesian, anti-individualists possess the intuitions that Solo is not self-conscious, and precisely because he exists in utter isolation. If you have the former intuition, then awareness of others is not a necessary condition of self-consciousness. However, if you have the latter intuition, then according to your conception of self-consciousness, awareness of others is a necessary condition of self-consciousness. But, volleying intuitions is a game that does not lead to profitable philosophy, and intuitions can hardly give us access to what is necessary or unnecessary, possible or impossible.

To ensure that we are not playing tennis with intuitions, I will suggest that we should focus on the necessary conditions of self-consciousness, what I call ‘the conditions strategy.’ The conditions strategy would begin with claims that express necessary conditions. For instance, “If a creature is to be self-conscious, then that creature needs to be aware of other creatures like itself.” This strategy would then not engage in thought experiments about our protagonist Solo, but instead would focus on the conditions in question and attempt to provide synthetic constructive arguments for those conditions. (I will discuss the nature of these synthetic constructive arguments in Chapter 1.) If those conditions can be maintained through these synthetic constructive arguments, then we are justified in inferring that Solo’s possession of self-consciousness is not so plausible, even if we are not willing to engage in a discussion about thought experiments.

The conception defended in these chapters may be described as an ‘interpersonal conception of self-consciousness’ or as ‘interpersonalism about self-consciousness.’ When I say ‘interpersonal,’ I want to claim that being presented with other persons is a necessary condition of being conscious of oneself as oneself. *The Symmetry Thesis* contends that insofar as one is capable of self-ascribing experiences to oneself, then one

must be able to self-ascribe experiences to other beings like oneself, like oneself in the respect of being a person. I will argue for what I will call ‘the Persons Theory’ The Persons Theory does not privilege the self or the other, but instead suggests that self and other are on the same level of priority. According to the Persons Theory, awareness of self and awareness of other develops in a single whirl of organism. It might be helpful to provide a taxonomy of the types of symmetry and asymmetry claims that are possible (this will become the focus of Chapter 6 on the type of subject sortal identification that is involved in self-ascription):

(1): *self-priority asymmetry*: the ability to self-ascribe is not dependent upon the ability to other-ascribe in the sense that one need not think about or perceive others in order to self-ascribe (in Chapter 6, I will discuss how the Simulation Theory of ascription of experiences (Goldman (2006)) is a self-priority asymmetry account);

(2): *other-priority asymmetry*: the ability to self-ascribe is not dependent upon the ability to other-ascribe in the sense that the ability to other-ascribe is prior to the ability to self-ascribe (in Chapter 6, I will discuss how the Theory-Theory of ascription of experiences (Gopnik and Meltzoff (1997)) is a other-priority asymmetry account of the ascription of experiences);

(3): *no-priority symmetry*: the ability to self-ascribe is dependent upon the ability to other-ascribe in the sense that one cannot self-ascribe unless one can other-ascribe and one cannot other-ascribe unless one can self-ascribe (in Chapter 6, I will discuss how the Persons Theory is a no-priority symmetry account of the ascription of experiences).

The Persons Theory about self-consciousness suggests that self-consciousness must be understood as fundamentally related to consciousness of other persons. The

strongest version of interpersonalism is that consciousness of other creatures is necessary for and constitutive of self-consciousness, such that a creature could not be self-conscious without awareness of others and awareness of others is what determines that one is self-conscious, because it is the crux of the development or genesis of self. As I have already suggested above, the opposing view would be an individualist conception of self-consciousness, or individualism about self-consciousness. To say that self-consciousness is interpersonal is to say that a creature cannot possess self-consciousness alone, that she needs to be aware of other creatures, but also that it is part of the nature of self-consciousness that it requires awareness of others.

2. Historical Arguments for the Symmetry Thesis

As I mentioned above, the purpose of the dissertation is to present a synthetic constructive argument for *the Symmetry Thesis*. Before I outline my argument in detail, I will provide some historical background for *the Symmetry Thesis*. Since the figures and arguments offered below would be difficult to discuss in their own terms, I will skate over interpretative issues and outline the arguments in broad strokes. Further, I do not critique these arguments here, but merely provide the arguments as a sketch of the types of arguments that serve as a background. Throughout the dissertation, I will make several points of contact with these historical arguments to ensure that my project has a proper historical grounding.

Generally, the most famous argument for *the Symmetry Thesis* is found in Hegel's (1977, pp. 104–138) *The Phenomenology of Spirit*. As far as distinctively philosophical arguments for an interpersonal conception of the self, of course, Hegel's account of the relation between self and other as characterized by the master-slave dialectic is

fundamentally a defense of interpersonalism. While I admire Hegel's holistic and systematic approach to metaphysics and philosophy of mind, for the purpose of inquiry into self-consciousness and *the Symmetry Thesis*, his approach makes it rather difficult to enter into a dialogue about the topic. Instead of Hegel, I will discuss Fichte as the central German Idealist figure to argue for *the Symmetry Thesis*. It could be argued that Hegel's account of self and other is already implicitly in Fichte's writing (Beiser (2002, pp. 341–3)). Most take the section entitled 'self-consciousness' in The Phenomenology of Spirit to be the first argument that self-consciousness depends upon awareness of others. However, J. G. Fichte's (2000) Foundations of Natural Right contains the first explicit argument for *the Symmetry Thesis*.

In that text, Fichte is attempting to provide an account of the conditions of self-consciousness. His account of self-consciousness is meant to achieve two basic purposes. On the one hand, Fichte argues that self-consciousness necessarily involves a distinction between "the I" and "the not-I." On the other hand, Fichte wants to preserve the pivotal relations between being a rational being and being a self-conscious being, namely that by being capable of reflection upon oneself thereby one is able to act in terms of laws and principles. For instance, if I were to act on the principle that I should only do to others what I can at the same time universalize as an action towards all others, I would have to have the capacity to reflect on my own actions. Thus, he suggests that by being a rational being, one is thereby capable of self-consciousness (an idea that was merely implicit in Kant's formulation of the categorical imperative). Fichte also follows Kant in thinking that self-consciousness is fundamentally an activity of the subject. What exactly is this self-activity? Self-activity is a constructive process that is governed by rules or

principles of synthesis. Kant suggests, ‘we can represent nothing as combined in the object without having previously combined it ourselves, and that among all representations **combination** is the only one that is not given through objects but can be executed only by the subject itself, since it is an act of its self-activity’ (1781, §B130).

Fichte takes this position farther and argues that self-consciousness a type of mental action, what he calls ‘self-activity.’ For Fichte, self-consciousness is the activity of a creature positing itself as self-positing, or positing itself as a being in the world. According to Fichte, the I posits itself as self-positing, by which he means that the individual of self-consciousness is conscious of itself as a self-determining individual. This is the first principle of Fichte’s philosophical system, that the individual is an activity of bringing itself into existence.

It follows from this view that there is not an agent prior to positing and there is not a product posterior to positing that is not itself that activity of self-synthesis. But, Fichte reasoned, the positing of “the I” cannot be understood as a purely creative force that brings about all experience for the individual, because the not-I exists independently as a material and sensible world, as Kant’s refutation of idealism (1781, §§ B274–294) had showed. Thus, it is a necessary condition of the self-positing I that there be a not-I, namely, the not-self or the material and sensible world. However, the argument continues, if the material and sensible not-I constrained “the I”, then the individual would lack all freedom and self-sufficiency, and this is implausible, given the centrality of Fichte’s first principle. Since “the I” is self-activity, then “the I” must maintain its freedom and self-sufficiency: it must be a subject that posits itself through the activity of synthesis, rather than being constrained by an external material world.

However, this provides Fichte with a dilemma. If the not-I is assumed to be a material or sensible force that constrains the I, then the I is determined (which goes against the first principle); however, if the freedom and self-sufficiency of the I is left unchecked, then the I will be allowed to create or produce the not-I, and hence become infinitely creative (which is absurd). Therefore, Fichte reasons, the concept of a self-positing individual requires the concept of the summons by other beings to become self-conscious, which must come from outside that self-positing individual (in the account in Chapter 6, this becomes the condition for what I call ‘recognition’). According to Fichte, the rational Other summons the individual to become self-conscious and at the same time asserts a check (in the account in Chapter 6, this becomes the condition for what I call ‘acknowledgement’) upon that individual to respect the rational Other in generating theoretical and practical rules and principles. The summons and check is what makes possible self-consciousness and can only arise assuming that the individual is aware of rational Others. Therefore, Fichte’s argument suggests that self-consciousness depends upon awareness of others, in this case awareness of other rational beings.

Another historical argument for *the Symmetry Thesis* can be found in the work of G. H. Mead, especially in his lectures Mind, Self and Society. According to Mead (1934), self-consciousness is an awareness of oneself as an object, however, it is an indirect object of experience. A creature can only be self-conscious by being aware of an object, but this object cannot be directly experienced. Mead suggests that self-consciousness is neither a reflection upon oneself in introspection, nor the perception of oneself as one’s body (This view will become pivotal to the discussion of the embodiment conditions of self-ascription in Chapter 2). Mead argues that the chief way

in which we can become aware of an object is through linguistic gestures called “the conversation of gestures.” For example, the chief way we become aware of a tree is through an individual’s referring to that tree.

Further, Mead conceives of the conversation of gestures as a rule-governed activity similar to a game, which cannot be played by one individual alone. One cannot play the game of the conversation of gestures and refer to a tree unless one recognizes that there are others that refer to trees and acknowledges the constraints upon that reference. This principle does not apply only to objects, but also applies to awareness of oneself as an object. One cannot be aware of oneself through a gesture towards oneself unless one is or has been aware of some other’s gesture towards oneself as that public object. According to Mead, “[the individual self] can enter as an object to himself only on the basis of social relations and interactions, only by means of his experiential transactions with other individuals in an organized social environment” (1934, p. 225). The individual is an object for others first and only through the linguistic interactions with others becomes a self-conscious individual. Therefore, Mead’s argument suggests that self-consciousness depends upon awareness of others, in this case awareness of the linguistic gestures of others towards oneself.

Another figure that argues for *the Symmetry Thesis* is Wittgenstein in his implicit arguments against individualism in Philosophical Investigations. In what has now become known as “the private language considerations,” Wittgenstein argues against (or at least manifests a dissatisfaction with) the individualistic notion of a language *qua* private language. His claim from the *Tractatus* that “the limits of language (of that which alone I understand) means the limits of *my* world” (5.62) was found to be deeply problematic,

because ‘my language’ is a term than cannot be given a sense. One could not have a language that was in principle merely one’s own, but instead one’s own language or idiolect is dependent upon a public language and the form of life in which that language is used. (Stephan Mulhall’s (2007) treatment of Wittgenstein’s private language considerations is a good recent discussion.) This could be taken to suggest, that similarly, one could not talk or think about oneself as oneself— whether this is understood as reference with “I” or in other terms— unless one could also talk or think about other self-conscious beings— whether this is understood as reference with “you” or in other terms. However, Wittgenstein’s interpersonal conclusion about language is difficult to extend to an argument for an interpersonal conclusion about self-consciousness, especially since Wittgenstein himself at times seemed to support individualism about self-consciousness ((1999, § 62-5.63); (1958, p. 61)). I will discuss this in Chapter 2 when I discuss the constancy of sense thesis about experiential predicates.

Following Wittgenstein’s lead, P. F. Strawson (1959) discusses self-ascription explicitly with the following claim: “it is a necessary condition of one’s ascribing states of consciousness, experiences to oneself, in the way one does, that one should also ascribe them, or be prepared to ascribe them, to others who are not oneself” (1959, p. 94). This glimpse of a thesis and argument occurs in the context of P. F. Strawson’s defense of the idea that the concept of a person is an indispensable concept (“a major logical type or categories of individuals”) for any descriptive ontology that admits states of consciousness. So, although P. F. Strawson provides a methodological framework for an argument against individualism about self-consciousness, he can only provide details

with respect to the acceptance of the argument about persons. But, even with such a limited scope the argument is revealing.

According to Strawson, self-consciousness is the self-ascription of experiential predicates to oneself. For example, if Solo were to be self-conscious, then he would need to be able to ascribe the predicate of having a headache. Strawson argues that in order to be able to ascribe the predicate of having a headache, one needs to master the application of that predicate. Mastery of the predicate requires that the predicate have generality and unity in its application, which means that Strawson's account is a criteriological account. (This criteriological account will be discussed in more detail in Chapter 5.) According to Strawson, one cannot master the application of that predicate unless one is capable of applying the predicate of having a headache to another individual. Therefore, in order for one to be able to ascribe predicates to oneself one must be able to ascribe predicates to others. Strawson's argument suggests that self-consciousness depends upon awareness of others, in this case the awareness of others possessing the predicates that are ascribed in self-consciousness as self-ascription. Further, the ascription of experiential predicates to oneself as a subject cannot be conceived to occur unless we understand such ascription in the context of the practice of ascribing mental state predicates to other subjects construed as persons, human beings, or people that take up the same form of life as the subject.

Strawson's argument is meant to oppose individualism about self-consciousness—that a self-conscious subject does not need to rely on other individuals to be conscious of itself as itself or its mental states as its own. Strawson argues that there is a central type of interpersonal relation that mediates the relation between a self-conscious individual and its awareness of other subjects, namely, the concept of a person.

So, to the extent that the other subject is a person, it is argued that when one is aware of a person, then one is aware of another subject. However, it should not be supposed that the relation between self and other should be understood solely as a relation between persons, between individuals as human beings and other subjects as other human beings. As I will suggest in these chapters, it is not only persons or human beings that possess self-consciousness. So, P. F. Strawson's arguments appear to support versions of *the Symmetry Thesis*, but upon further interpretation (which I will articulate throughout the dissertation), are either short on defense of key premises or are actually premises in arguments for conclusions that have little to do with the nature of self-consciousness. Given that *the Symmetry Thesis* should be adjusted to reflect biological inclusiveness about the concept of self-consciousness, it might be phrased as: if a creature is capable of self-ascription, and self-ascription is in its nature interpersonal, then that creature will be required to have been aware of other creatures like itself, or to use the general term, to be aware of that creature's conspecifics. (As we will see in Chapter 6 and the conclusion, *the Symmetry Thesis* is also supported by recent evidence in psychology and neuroscience.)

I have not criticized the truth of the premises or the logic of these arguments, since I will return to such arguments for *the Symmetry Thesis* throughout the dissertation. Instead, I want to point out some basic features of these arguments, in order to highlight the basic introductory claims that need to be made before proceeding. There are some similarities between these historical arguments for *the Symmetry Thesis*.

First, each argument employs a methodology that is resoundingly Kantian (even if not always explicit in Wittgenstein, Wittgenstein's reading of Kant's First Critique was

significant for his work) in that even when they are not interpreting Kant, they adapt Kantian views, concepts and arguments for introduction into the discussion. As I mentioned above, although Kant is often considered an individualist about self-consciousness, the seeds for interpersonalism are implicit in various texts of his later critical philosophy, although this will not be the focus of the dissertation.

Second, the methodology is taken up with employing transcendental arguments that instruct us about the major facets of our understanding, where our understanding is an abstraction from our theoretical assumptions, where those assumptions exist on a continuum from the philosophers' colloquium to the everyman's bodega. A transcendental argument is an argument that begins with an experiential fact or state of affairs, and elucidates the necessary conditions of that fact or state of affairs. Once it is argued that there are necessary conditions of a fact or state of affairs, the transcendental argument concludes that the conditions need to obtain. In the next chapter, I will discuss the use of transcendental arguments in the philosophy of mind.

Third, they appreciate the contribution of the historical tradition of philosophy, borrowing from this tradition to help illuminate our understanding of the contemporary problems that we grapple with. Finally, the methodology is committed to naturalistic empiricism in the broad sense that involves the commitment to the idea that what exists obtains in the natural world, which is revealed predominantly through sense experience. While I will accept this methodology in these chapters, I want to at least begin to analyze the argument provided by P. F. Strawson's assessment of self-consciousness and *the Symmetry Thesis*.

3. The Constructive Argument for the Symmetry Thesis

My purpose is to provide a thorough defense of *the Symmetry Thesis* through the argument I summarize below. This argument is loosely inspired by Bermúdez’s discussion of *the Symmetry Thesis* in The Paradox of Self-Consciousness. As will become clear in the next chapter, I use Bermúdez’s interpretation as a foil to develop my own constructive argument. As I mentioned above, I stipulate the kind of self-consciousness that I am concerned with: self-consciousness is the self-ascription of experiential predicates. The self-ascription of experiential predicates has a subject-predicate form. For instance, “I am hungry” ascribes the experiential state of being hungry to a subject. In Chapter 2, I argue that because of its structure, self-ascription has to meet a constraint called ‘the Generality Constraint.’

Gareth Evans (1982) defines the generality constraint as: if S is to think “a is F” then S would need to meet the following conditions: (GC₁): S would need to be able to think “b is F,”; (GC₂): S would need to be able to think “a is G” (for any concept G that S possesses which is not F). I do not argue for the generality constraint or compositionality in general; instead, the argument is directly specifically about self-ascription. The basic argument from the generality constraint is that if a subject is to be able to self-ascribe the experience “I am F,” then she must have two distinct abilities: (GC₁) she must be able to ascribe experiential predicates, e.g., “b is F,” “c is F,” to arbitrarily distinguishable individuals; (GC₂) she must be able to ascribe “I am G,” “I am H,” for any experiential predicates she possesses independent of “I am F.”

If self-ascription requires meeting the generality constraint, then this requires the ability to generalize the subject and the ability to generalize the predicate involved in self-ascription. In Chapter 2, I will distinguish between those theses that follow from

generalizing the predicate and those that follow from generalizing the subject. There is already an extensive literature on the topic of generalizing the predicate of self-ascription, mostly following up on Ludwig Wittgenstein's claim that an individual could not apply a predicate to herself only, as I said, she could not have a private language (1958).

However, my focus will be on generalizing the subject of self-ascription.

I argue that generalizing the subject of self-ascription depends crucially on identification of the subject of experiential predicates. In Chapters 3 and 4, I admit that this requires dismissing the idea that self-ascription is immune to error through misidentification (IEM). In Chapter 3, I argue that IEM depends pivotally on the assumption that self-ascription comes in two varieties, ascription to oneself as a subject and ascription to oneself as an object. I argue that there is no principled distinction between the two that can serve as the basis for IEM, which provides further support for the account of self-ascription as dependent on the generality constraint that I presented in Chapter 2. In Chapter 4, I present two arguments against the view that self-ascription is IEM: one conceptual (drawing on the work of Gareth Evans (1982)) and one empirical. In this respect, then, self-ascription is not IEM, so I am free to conclude that self-ascription depends upon a type of self-identification.

Once I have argued that self-ascription depends significantly on self-identification, I argue in Chapter 5 that such identification cannot be done in terms of experiential predicates alone, but instead in terms of the subject of those predicates. In this respect, I part ways with Bermúdez (1998, p. 241), who thinks that self-identification is achieved through a piecemeal description of the subject through entertaining experiences: perceptions, actions and emotions. Independently of my work on

Bermúdez’s arguments, I have found that much of the discussion of Strawson and Evans’ argument for *the Symmetry Thesis* involves the tendency to adhere to a problematic principle of self-ascription, what I call the ‘predicate principle,’ which maintains that in order to self-ascribe, it is sufficient for one to identify the subject of self-ascription through the content of the experiential predicates of self-ascription.

In Chapter 5, I argue for a different principle of self-ascription— what I call ‘the Persons Principle’— that puts Strawson and Evans’ account of self-ascription in the proper light. The key principle I defend is the following: in order to account for self-ascription, one must identify the subject of self-ascription independently of the experiential predicates of self-ascription.

Once I make clear that the subject of self-ascription needs to be identified in independent terms, in Chapter 6, I argue that while there are many ways by which a subject may identify itself, the central way is by way of the recognition and acknowledgement of persons. In order to identify oneself as a subject one must identify oneself as a type of being. In the human case, one must identify oneself as a person. I argue that if it can be shown that creatures need to have the ability for intraspecies recognition and acknowledgement in order to identify the subject, then, one must be able to recognize and acknowledge the form of life in which one participates. I focus on the phenomenon of joint engagement to bolster my argument for the Persons Theory.

In the conclusion, I discuss the import of my argument for the Symmetry Thesis. I argue that *the Symmetry Thesis* enables us to overcome the conceptual problem of other minds. In the next chapter, I turn to the justification of my methodological assumptions.

Chapter 1

“The task facing developmental psychologists is not that of mapping the kinds of evidence on the basis of which infants attribute mental states to others, but that of giving an account of the way in which infants’ perceptual experience becomes imbued with psychological concepts. It is here that social interaction may have a constitutive role to play”
— Johannes Roessler (2005, p. 252).

The I is a We; the We, an I: Constitutive Claims, Development and Transcendental Arguments

As I discussed in the introduction, there are differences between disciplines about the import of *the Symmetry Thesis*. Some think that the nature of self-consciousness is constituted by relations to other individuals like ourselves; while others think that the nature of self-consciousness is not so constituted. According to this latter position, self-consciousness develops independently of the influence of other creatures. The typical approach to consciousness and self-consciousness in the discipline of philosophy is by way of the latter in that it greatly ignores the relationship between the self and other. As I discussed in the introduction, the tendency in philosophy of mind is to presuppose individualism about self-consciousness. However, the common approach in psychology, especially in developmental psychology, takes the relationship between self and other to be developmentally intertwined in a way that makes the development of self-consciousness constitutively depend interactions between selves and others.

The goal of this chapter is to consider this contemporary dialectic through a discussion of what I will call ‘the constitutive claim’ concerning the self-ascription of experiences. According to one recent philosophical psychologist, J. L. Bermúdez, there

is “a constitutive link” (1998, p. 247) between awareness of one’s own mind and awareness of other minds. The synthetic constructive argument for *the Symmetry Thesis* that I laid out in the introduction required defending this type of constitutive claim: *the ability to ascribe experiences to oneself requires the ability to ascribe experiences to other subjects like oneself.*

I will argue that the requirement that is expressed by this constitutive claim concerns a necessary condition of self-ascription. However, Bermúdez (and others) do not recognize that defending the constitutive claim requires a transcendental argument and explicitly deny the need for an argument for a necessary condition. Unfortunately, this is partly because transcendental arguments have an undeservedly bad reputation. The aim of this chapter is to argue for the importance of transcendental arguments in the philosophy of self-consciousness, in particular with respect to the constitutive claim. As will become clear below, I am not arguing that a purely a priori approach to self-consciousness is desirable. In fact, I think that transcendental arguments are synthetic a priori arguments, but the a prioricity is no more divorced from experience than are arguments in arithmetic and geometry.

In the introduction, I discussed a few predecessor arguments that supported something like *the Symmetry Thesis*. There are three points that are distinctive about those arguments. First, they each begin with the question of how a certain phenomenon of self-consciousness is possible. For example, in general terms, how is it possible for an organism to think about itself as itself? This question is what I will call a ‘how-possible question’ about self-consciousness. For example, in Fichte’s argument, he considers how the self-activity of the rational agent is possible. Second, they engage in an initial

response to a how-possible question by offering a developmental account, because the phenomena suggest that certain means are involved in accounting for it. In the initial response, the focus of the explanation is on the process of development, but as will become clear below, the concept of development is multiply ambiguous. For example, Fichte considers two genetic means: on the one hand, the subject might be self-constituting; on the other hand, the material world might be what constitutes the subject. But, both of these arguments by way of means are found to be problematic. Instead, the third point is to provide a transcendental argument for a constitutive claim that a certain necessary condition is responsible for the development of the phenomenon in question, which builds on the initial response, but transforms that response into a claim about a necessary condition. For example, for our purposes here, Fichte considers the recognition and acknowledgement of other rational beings (among other conditions) as constitutive of one's self-consciousness. The point of this chapter is to defend the idea that to say that interpersonal interaction is 'constitutive' of self-ascription is to say something more than that it is developmentally prior (As I will make clear below, this three step procedure of the transcendental argument is replaced by Cassam's (2005) account of transcendental argument in his book The Possibility of Knowledge).

As I briefly discussed in the introduction, the goal of the dissertation is not to focus on historical figures nor is it to focus merely on transcendental arguments. My strategy is unique in that I am arguing that if we can legitimate a type of transcendental psychology about self-consciousness inspired by Kant (though one that he himself failed to engage in the First Critique), then *the Symmetry Thesis* becomes a thesis that both accounts for the necessary condition of self-ascription, while remaining beholden to its *a*

posteriori or empirical conditions. This might seem puzzling to some. One reason that one might be puzzled is that there is a tendency for philosophers that use transcendental arguments to ignore empirical findings. One clear example of this is Cassam's book Self and World (Andrew Brook (1997) argues this in his review of Cassam's book.). I do not think there is a principled reason why investigation by transcendental argument—and in general a transcendental psychology concerning self-consciousness—is committed to ignoring findings in empirical psychology. Another reason that one might be puzzled is that philosophical psychologists tend to err on the side of the *a posteriori* or empirical conditions by merely pointing at the evidence as a bare Given. Psychologists that delve into the dependency relations between self-ascription and other-ascription tend to minimize the need for transcendental arguments and the defense of necessary conditions, but this merely cooks the books in favor theoretical assumptions. The assumption in those cases is that if X is developmentally prior to Y, then X constitutes Y. "For what other sort of constitution could be involved other than causal-cum-developmental causation?" they ask. My methodological argument here is that we should bring these two traditions together by adopting the strengths and leaving aside the weaknesses of each. One merit of my project is that I bring the transcendental arguments of P. F. Strawson and Gareth Evans into contact with research in the cognitive sciences. The main point of this chapter is that transcendental arguments need not be at odds with empirical concerns and that, especially in the context of discussion of self-consciousness, transcendental arguments deserve a deeper consideration than they currently get.

My work on *the Symmetry Thesis* is also relevant for developmental psychology, especially the work of James J. Gibson (1986) and Ulric Neisser (1993) on the

development of the self. The project enables me to engage with the recent interdisciplinary focus on what I call ‘joint engagement’ (As I will discuss in further chapters, ‘joint engagement’ refers to the activities of jointly perceiving, acting and emoting.). The normal human experience involves early and often engagement with others and the environment via perception, action, emotion and other mental states. The desert island upbringing of Solo that was discussed in the introduction is extremely rare, and may even be impossible. Our normal course of development involves engaging with each other and sharing perceptions, actions and emotions in a way that significantly informs our understanding of ourselves. It is for this reason that joint engagement is pivotal for our understanding of the relation between self and other. Joint engagement includes: joint perception (for instance, an infant and adult jointly attending to an object like a burning candle); joint action (for instance, infant and adult engaging in parallel actions like imitation in peek-a-boo); and, joint affection (for instance, infant and adult entertaining emotions that require intersubjective relations like pride and shame). Philosophers and psychologists have considered how two individuals share perceptions, actions and emotions (Neisser (1993); Moore and Dunham (1995); Eilan, Hoerl, McCormack, and Roessler (2005); Rochat (2009), and Seemann (2010)). I have incorporated work on joint engagement in a way that inspires a revision of the optimistic attitude towards individualistic notions of self-consciousness. In this way, my approach to these joint phenomena enables a discussion between developmental psychology and traditional topics in the philosophy of mind.

Again, the issue then is how to build a bridge between these two approaches, how to build a bridge between the strictly philosophical approach and the strictly

psychological approach to the question about the relationship between consciousness of oneself and consciousness of others. Once we consider the grounds or conditions of self-ascription, it might be helpful to discuss the different approaches as *a priori* approaches or *a posteriori* approaches. An *a priori* approach would reflect upon the most abstract and general conditions for self-ascription. As Kantians are fond of writing, it is concerned with experience in general, meaning the universal and necessary features of experience. An *a posteriori* approach would focus on a more concrete and specific type or kind of self-ascription, or it is concerned with experiences in isolation from the general case. For example, *a posteriori* approaches might focus on the development of human self-ascription by considering its ontogenetic or phylogenetic origins. One might think that articulating self-ascription via the *a priori* approach would not require any discussion of the *a posteriori*. Alternatively, one might think that considering the *a posteriori* findings or results wouldn't require any *a priori* reasoning. Let me turn to some issues with the purely *a posteriori* approach.

It might be objected that the kind of methodology I am using does not allow for an *a posteriori* approach to inquiry into self-consciousness. One might use a strategy for a defense of *the Symmetry Thesis* that would be a more *a posteriori* approach, which would aim to discover the nature of self-consciousness in human and non-human animals through observation of actual cases and experiments. One kind of *a posteriori* methodology to defend *the Symmetry Thesis* might be to merely point to the simple fact that other psychological subjects exist, other human beings exist, other language users exist. There is no doubt a kind of traction to be gained by making a parallel claim to Johnson's attempt to refute Berkeley's immaterialism by kicking a stone. (At a recent

interdisciplinary conference about joint engagement, a prominent developmental psychologist said, “But, you and I both know there is no problem of other minds.”) We might follow Johnson’s lead and simply appeal to the fact that self-conscious creatures are as a matter of natural fact surrounded by others, whether as human beings are surrounded by other persons, or monkeys are surrounded by other primates. The suggestion here is that in order to argue for the interpersonal conception of self-consciousness, all that would be required is to argue for an interpersonal conception of the individual, according to which the individual (either by definition or by nature) is an interpersonal entity. One way of understanding the claim that a self is an interpersonal entity is to say that the conception of an individual self (either by definition or by nature) is intrinsically related to the conception of others.

While this view may seem appealing, it is not the strategy that I will adopt as a starting point. I think we have committed great errors by beginning with the concept of the self or the concept of the individual and thereupon making conclusions about self-consciousness. I will not address the question of what a self is. (For a different attitude towards the priority of the metaphysics of selves over the nature of self-consciousness, see G. Strawson (2009)). Presumably, in order for us to approach this question, we have to ask, “What do we experience as a self?” But, we cannot answer this question unless we assume that our experience is self-conscious experience, so the question, “What is the nature of self-consciousness?” is prior to any questions probing the phenomenology of self-experience. We might discover that it is in the nature what a human self is that it is intrinsically interpersonally related, but even if this is true, it does not follow from this that self-ascription of experiences depends upon interaction between self and other.

Interestingly, in normal introspective self-consciousness, we are presented to ourselves in a way that does not require being aware of other subjects. Alternatively, we are not aware of the intrinsic interpersonal relation between self and other, but from this it does not follow that selves are not in fact so related to others. We should not necessarily infer from the way things happen to be to the way things must be, neither should we necessarily infer from the way things seem, appear or are presented to the way things are. But, this abstract constraint surely puts us in a dismal predicament.

To bring us out of that impasse and to bring this abstract lesson down to earth, I suggest the following. Although we may observe that awareness of ourselves as living animals is as a matter of fact socially determined— for instance, humans, primates and other animals, for instance, some social birds like geese— become self-conscious while in the company of other subjects of consciousness, this need not be the way it is necessarily. It is an open question whether organisms might develop self-consciousness in isolation. Further, if we seem, appear, or are presented to ourselves as ourselves in a way that suggests that we could, would, or should be capable of self-consciousness in the absence of others, this need not allow us to infer that the nature of self-consciousness is individualistic. So, it is not sufficient to argue for the interpersonal conception of self-consciousness that one appears to be aware of a self that as a matter of fact is an interpersonal entity. It may be that the self is an entity that has a social upbringing among other persons, other agents, or other rational beings. However, even if it happens to be this way for humans or other animals, it is not the case that the conception of self-consciousness needs to be understood in this way.

It might be objected that this means that I am suggesting that there is a metanatural or metaphysical necessary connection between selves and others. (I discuss such an objection below as the modal objection to transcendental arguments.) Instead, I am merely suggesting that although our experience may present us with certain facts about the human or even non-human individuals, given our intellectual situation *vis-à-vis* self-consciousness, we need to allow a respectable distance between our theoretical claims about the content of self-consciousness and our commonsense and scientific observations of human and non-human individuals. The methodology involves engaging in a reflective equilibrium between our theoretical claims about biological creatures' experience, on the one hand, and our everyday observations of biological creatures, on the other hand. I offer two reasons to adopt such a transcendental perspective about self-consciousness.

The first reason is that when we account for the development of self-consciousness, we do not usually stray from the observation of normal human narratives about the development of self-consciousness. When we discuss human self-consciousness developmentally, interpersonal relations are a central part of the narrative. Mostly, however, awareness of other subjects is represented in those narratives as a stage or phase in the genesis of self-consciousness. For instance, we might study joint attention, social referencing and imitation in the first 3 years of human infancy, and discover that interpersonal relations are pivotal in the genesis and development of self-consciousness. However, this approach does not have import for whether interpersonal relations are necessary conditions, because this approach confuses a developmental or genetic condition— a condition that is necessary for the origin of X— with an

ontological condition— a condition that is necessary for the existence of X. As I will make clear below in my critique of Bermúdez’s account of joint engagement, one cannot wring a constitutive claim out of a merely correlational or causal claim.

The second reason is because the central issue about whether others are necessary for self-consciousness makes it difficult to articulate in a non-question begging way what counts as normal self-consciousness and what counts as abnormal or disordered cases of self-consciousness. When we observe non-normal or disordered cases of self-consciousness, for instance, autism or schizophrenia or marginal cases of minimal interpersonal relations such as isolated or feral children, we still are left with questions about whether the individuals suffering these conditions are self-conscious. Is it proper to attribute self-consciousness to autistics, schizophrenics and isolated/feral children? I do not doubt that we may have intuitions that autistic individuals possess deficient self-consciousness despite lacking interpersonal relations, or that schizophrenic individuals do possess excessive self-consciousness possibly because of an increase in imaginary interpersonal relations. (Again, I ultimately will suggest in the conclusion that my prediction is that once we gain clarity and precision about self-consciousness, then the phenomena concerning autism and schizophrenia ultimately support *the Symmetry Thesis*, but this conclusion should be taken as a conservative hypothesis to be researched by empirical psychologists.)

However, without clarity and precision concerning the concept of self-ascription, we cannot ground these intuitions. It is not simply a matter of a lack of conceptual clarity and precision, there is also an experimental bottleneck that we face. While our well-founded intellectual curiosity about whether interpersonal relations are necessary for self-

consciousness might lead us to devise isolation and solitude experiments, we cannot follow *Psamtik I* and put a boy in a box just to see if he comes out self-conscious. This inquiry into the *a posteriori* methodology raises the questions, “What is the nature of self-ascription?” and “What are the necessary and sufficient conditions of self-ascription?” I will call these questions “the conditions questions”. The conditions questions provide limits to the types of arguments we can provide for *the Symmetry Thesis*. It should become clear that I am not concerned with mediating between the *a priori* and the *a posteriori* approach. I consider myself more of a philosophical playground moderator. The best way is likely through the middle way, which I will attempt to spell out here.

I will say a bit more about the integrative methodology of self-consciousness that I apply in the dissertation. My methodology is integrative in that it brings together *a priori* and *a posteriori* approaches. Such a methodology considers self-consciousness on some occasions, from the armchair and on other occasions, from the lab. I am suggesting that we should follow Sellars’ conception of a ‘stereoscopic vision’ (Sellars, 1963, p. 4) and integrate the manifest image of self-consciousness as disclosed by analysis of our conceptual scheme with the scientific image of self-consciousness as discovered by research in psychology and neuroscience. The integrative methodology then integrates results from reflection on necessary and sufficient conditions, on one end of the spectrum, and results from observation of genesis, development, and the inner workings of the brain, on the other end of the spectrum.

One might object that this methodology is anti-naturalistic. However, my methodology is natural in that it considers the human being that is self-conscious as part

and parcel of the natural world. As I will discuss in Chapter 5, my work on self-consciousness and the self is inspired by James J. Gibson and Marjorie Grene's work, both of which hold biological or ecological accounts of the subject. As will become clear in Chapter 6 on joint engagement, my notion of the subject is as a person taken up in a biological form of life— a person among persons.

My methodology is phenomenological in that it considers the phenomenon of self-consciousness to be central to inquiry. Unlike traditional phenomenology (such as that inspired by Husserl), however, my aim is to set aside the privileging of introspection and the first person perspective. Some may say that we cannot do phenomenology except by privileging introspection and the first person perspective. But, this reflective notion of phenomenology need not be the only means to understanding the phenomena— a proper phenomenology looks at subjectivity from first person, third person, and importantly for our purposes, second person perspectives. The major negative critique of the dissertation is that the second-person perspective is lacking in our study of consciousness and self-consciousness. Going back to the phenomena does not mean focusing on what it's like to experience such and such (*pace* Thomas Nagel (1986)), for that is a limited notion of phenomenology. I now turn to an elucidation of the central claim of this chapter— the constitution claim about self-ascription. However, before I do, I need to provide some background concerning Bermúdez's motivation for making this claim.

1. Bermúdez's Constitution Claim

José Luis Bermúdez's constitution claim occurs in the context of his larger project concerning self-consciousness. My focus on Bermúdez here is not out of figure worship; nor is my project merely a matter of heedless critique. I focus on Bermúdez as a foil,

because while *the Symmetry Thesis* has a long history, Bermúdez (1998) is responsible for coining its name. My project takes the recent work of Bermúdez as a springboard for the following reasons: (1) His work is exemplary because he offers one of the only systematic accounts of the nature of self-consciousness in recent philosophy and psychology; (2) Bermúdez exemplifies the kind of interdisciplinary work in philosophical psychology that is valuable to both disciplines; (3) Bermúdez locates the origin of *the Symmetry Thesis* in the work of P. F. Strawson (1959) and Gareth Evans (1982), two figures that epitomize the kind of approach to self-consciousness that I will adopt. Bermúdez’s approach and my approach are different in many respects as will become clear throughout the dissertation. But, the central way that we differ is that Bermúdez thinks that the Strawsonian defense of *the Symmetry Thesis* is inevitably unsuccessful. However, I disagree, and I think much more can and should be said in its favor. Before I begin to focus on Bermúdez’s constitution claim, I need to provide some background concerning the putative paradox of self-consciousness.

Much of the philosophical literature on self-consciousness focuses on self-reference, specifically on epistemic or semantic peculiarities of self-reference (Perry (1979); Casteneda (1999); Shoemaker (1968)). It is usually assumed that self-ascriptions are incorrigible (Mackie (1963)), meaning they cannot be corrected. If I say or think about my own mental state, for instance, “I have a headache,” then I cannot be incorrect about my having a headache. It is also usually assumed that self-ascriptions are immune to error through misidentification. If I think, “I see a garnet ring”, then I may be wrong about it being a garnet ring and even about it being an episode of seeing (because for instance, I may be imagining it). However, I cannot be wrong about who is seeing,

namely myself. Philosophers tend to assume that if we focus solely on self-reference—specifically, on utterances employing the first person singular pronoun ‘I’ like “I have a headache”—then this will allow us to better understand the nature of self-consciousness. However, it is arguable that individuals—both human and non-human—that are not capable of using ‘I’ (because they are pre-linguistic or non-linguistic animals) nevertheless possess self-consciousness. For this and other reasons, philosophers and psychologists have begun to buck the trend that focuses on self-reference in order to understand self-consciousness.

As a leading philosopher of self-consciousness, Uriah Kriegel, suggests when he writes: “current work on self-consciousness does not appear to address the need for a general theory thereof. Instead, it rests content with a piecemeal treatment of each alleged [epistemic and semantic peculiarities] in separation from the rest. Sooner or later, however, this will have to be rectified by a reorientation or reorganization of research in this area” (2008). Kriegel suggests that the work of José Luis Bermúdez is an important contribution to reorientation of research, because Bermúdez provides a rich and interesting account of the nature of self-consciousness through its basis on more basic cognitive abilities. My work is very much inspired by this reorganization of research. As will become clear in chapters 2–4, I will try to bring about a reorientation and reorganization of research into self-consciousness by providing a functional account of self-ascription, questioning the subject/object distinction, and critiquing the tendency to rely on IEM as a mark of self-consciousness. However, before I provide my own account, I need to provide some background concerning Bermúdez’s work.

In his pivotal work, The Paradox of Self-Consciousness, Bermúdez motivates an account of self-consciousness through the elucidation of a paradox. He argues that any attempt to account for self-consciousness in terms of self-reference will run into a paradox: self-reference must be explained in terms of I-thoughts; I-thoughts must be explained in terms of self-reference; therefore, we run into circularity in attempting to account for self-consciousness. (To lay my cards on the table in advance, my goal in the dissertation is to show how self-consciousness is a type of self-ascription, that self-ascription depends upon I-thoughts; but that I-thoughts can be explained in functional terms *via* description of linguistic and non-linguistic behaviors that do not require appeal to self-reference.) Bermúdez lays out the paradox with the following propositions:

- (1) The only way to analyze what is distinctive about self-consciousness is by analyzing the capacity to think ‘I’-thoughts;
- (2) The only way to analyze the capacity to think a particular range of thoughts is by analyzing the capacity for the canonical linguistic expression of those thoughts (the Thought-Language Principle);
- (3) ‘I’-thoughts are canonically expressed by means of the first-person [singular] pronoun (FPSP);
- (4) Mastery of the first-person [singular] pronoun requires the capacity to think ‘I’-thoughts;
- (5) A non-circular account of self-consciousness is possible;
- (6) Mastery of the semantics of the first-person pronoun meets the Acquisition Constraint.

Bermúdez rejects premise 2 to resolve the paradox. He offers an account of self-consciousness that does not analyze self-consciousness in terms of self-reference. Instead, Bermúdez argues that self-consciousness needs to be analyzed and explained independently of the analysis of utterances employing the first person singular pronoun ‘I.’ He argues that there are real levels of self-consciousness that are non-conceptual—ecological self-awareness, proprioceptive awareness, awareness of one’s embodied agency, an experiential point of view, and ‘psychological self-awareness’—none of which require concept mastery *qua* language mastery. According to Bermúdez, the account of self-consciousness couched in terms of more basic phenomena allows philosophers and psychologists to understand self-consciousness as conceptually distinct and developmentally prior to self-reference, which resolves the paradox of self-consciousness. While I agree with much of this strategy, I offer two reasons why I will not sign on to the approach via the alleged paradox of self-consciousness.

The first reason is that unless one accepts the assumptions of the paradox, then one is not forced to accept the conclusions of the account of self-consciousness. The premises of the paradox may simply be unconvincing. So, in this respect, while Bermúdez has offered a way to avoid the piecemeal analysis of epistemic and semantic peculiarities of self-consciousness (and therefore offers a way to provide a systematic approach to self-consciousness), he has merely supplanted this approach with the spotlighting of a potentially inert paradox. I say it is inert because one man’s *modus tollens* is another man’s *modus ponens*. Bermúdez wants to reject premise 2 in order to obviate the paradox. However, one could easily accept premise 2 and argue that the types of self-consciousness that Bermúdez offers are not really forms of self-

consciousness at all, since he has provided no argument that they count as legitimate forms of self-consciousness. (This critique was inspired by a conversation with Kristina Musholt.)

The second reason that I will not undertake inquiry into self-consciousness via the paradox of self-consciousness is that what Bermúdez suggests is a paradox of self-consciousness is really a paradox of self-reference, given that self-reference is the target explanandum. He assumes that self-consciousness is self-reference and needs to be explained. But, then he argues that self-consciousness should not be understood in terms of self-reference, but in other terms, namely, in terms of levels of self-awareness that do not require self-reference. That suggests that self-consciousness is both self-reference and is not self-reference. However, it is obvious that Bermúdez would not want to subscribe to such an outright contradiction, so for this reason it is still unclear what self-consciousness is, except that it is a multifarious concept that applies to many different levels of self-consciousness phenomena.

Thus, instead of focusing on the supposed paradox, I will focus on the level of self-consciousness that Bermúdez discusses in Chapter 9— “the explicitly psychological dimension of self-awareness” (1998, p. 229), given that this level is central to most accounts of the nature of self-consciousness as self-ascription that I mentioned in the introduction. I will articulate this explicitly psychological dimension of self-awareness as self-ascription through a functional account of I-thoughts. I will say more about self-ascription in the next chapter. Given that I agree with much of Bermúdez’s project— both the need to present a general account of self-consciousness and the importance of

the empirical study of more primitive forms of self-consciousness— I will not dispute his analysis of self-consciousness in terms of different levels.

One need not appeal to a paradox to motivate the description of self-consciousness in terms of levels. One simply way is to point out that ‘self-consciousness’ is said in many ways. We use ‘self-consciousness’ to refer to self-reference with the first person singular pronoun, to self-ascription of experiences, to the functional unity of our experiences, to the self-presentation of our bodies to ourselves, and to self-discrimination, the discrimination of ourselves from the world. These different concepts of self-consciousness seem to elucidate different levels of self-consciousness. But, in the dissertation, the discussion of self-consciousness as self-ascription will take center stage. I agree that self-ascription surely has empirical conditions and Bermúdez is correct that empirical conditions are central. It will become clear below that my version of these levels of self-consciousness is best understood as the empirical conditions of self-ascription. However, I will focus on his account of “psychological self-awareness” because I think there are interesting philosophical and psychological issues that need to be addressed to properly account for this level of self-consciousness. So, what exactly is psychological self-awareness? Bermúdez assumes that “self-ascription of psychological states in the subject-predicate form” (1998, p. 233) is the paradigm case of psychological self-awareness. While it is not clear what paradigm case means here, it is clear that most consider self-consciousness to be best understood in terms of self-ascription.

For instance, P. F. Strawson (1966, p. 94ff) argues that self-consciousness is best understood in this way. Evans (1982, p. Chapter 7) focuses on self-ascription to

illuminate self-consciousness. Campbell (1994) argues that self-consciousness is best understood in terms of self-ascription of experiences. Cassam (1997, p. 118) argues: “a self-conscious subject must be capable of thinking of her experiences as her experiences, that is, of self-ascribing them.” About this paradigm form of self-consciousness, Bermúdez’s asks the question, “How is self-consciousness possible?” (1). He answers the how-possible question about self-ascription by suggesting that there might be an argument *the Symmetry Thesis*. He argues that self-ascription is “constituted” by the ability to apply psychological states in the subject-predicate form to other psychological subjects.

According to Bermúdez, there is “a constitutive link” between first-personal self-ascription and second- and/or third-personal ascription. On this view, if one is capable of ascribing psychological predicates to oneself then one must be capable of ascribing psychological predicates to other subjects. According to Bermúdez, psychological predicates include (but are not limited to) predicates of perception, for example, “I see a tree,” action, for example, “I am trying to walk,” of emotion, for example, “I am ashamed” and of thought, for example, “I am thinking about the Highlanders of Elgol.” He understands other-ascription to involve either ascribing at least some of those psychological predicates to other subjects (the weak reading of *the Symmetry Thesis*) or ascribing all of those psychological predicates to other subjects (the strong reading of *the Symmetry Thesis*).

A few exegetical points need to be made here which relate to my aims in the dissertation. The first exegetical point is that if we ask, “What makes self-consciousness possible?” and self-consciousness is understood in terms of self-ascription, then

presumably a project like Bermúdez's would turn to how self-ascription depends upon more basic types of self-consciousness. Bermúdez might have attempted to answer the how-possible question by providing an articulation of how self-ascription depends upon more primitive phenomena— having a sensibility, possessing proprioception, entertaining a non-conceptual point of view, etc. As we will discuss below, this would involve articulating the means by which self-consciousness develops. The idea would be that what makes possible self-ascription of experiences are these more basic or primitive forms of self-consciousness. However, Bermúdez does not do this. He neither shows how if self-ascription is dependent upon the previous levels of self-consciousness, how it is so, nor if it is not dependent upon the previous levels, why the previous levels are not sufficient to make it possible. In the following, however, I will make sure to describe how I take self-ascription to depend upon previous levels, making clear and precise how those previous levels are not sufficient.

The second exegetical point is that Bermúdez's account of self-ascription creates an ambiguity in the notion of other-ascription, an ambiguity that is essential to clarify. He confuses the notions of second-person and third-person ascription, blurring the relations that hold between them. We will call this 'the Bermúdez triangle.' For example, "You have a headache" and "John has a headache" place very different demands upon the subject making such ascriptions. This is common ambiguity in discussions of the other, partly because it is still unclear what the entailment relations are between second-person and third-person ascription. Does third-person ascription depend upon second-person? Do I have to be able to address a subject with 'you' in order to be able to address that same subject with some third-person term, like a name or

description? Does second-person ascription depend upon third-person ascription? Do I have to be able to address a subject in at least some third-person terms in order to be able to address that same subject with 'you'? I will come to answers these questions in due course, but for now these questions suffice to highlight this pivotal issue. As I said above, it is pivotal to the proper account of self-consciousness that we consider the first, third, and second personal approaches to self-consciousness.

The third exegetical point is that as I mentioned above, Bermúdez's strategy is to look for *non-conceptual contents* below the level of self-reference in order to obviate the paradox of self-consciousness. I must part ways with him at this point as well. I will not enter into the conceptual/non-conceptual debate about whether the contents of self-consciousness are conceptual or non-conceptual. I do not think that anything very fruitful will come from that debate about self-consciousness. I have two reasons for thinking that the conceptual/non-conceptual debates about self-consciousness are barren.

The first reason is that the principal whipping boy for non-conceptualists is John McDowell's (1996) conceptualist position about perception, action, thought and the self that he discusses in Mind and World. However, his theory or account of concepts is not clear or precise enough to evaluate. As far as I can tell, McDowell never offers an account of *the nature of concepts*, partly because of his anti-constructivist leanings. Given this lacuna, it is difficult to find non-conceptualists offering a proper interpretation of McDowell's position on the nature of concepts. Apart from this difficulty with McDowell interpretation, there is little consensus about what concepts are, so it has become increasingly difficult to determine how the negatively defined 'non-conceptual' gains a meaning.

The second reason is, apart from a lack of intelligibility about what concepts are— whether in the work of McDowell or in the discipline in general— the term ‘non-conceptual’ is negatively defined (Gunther (2003)). So, a negatively defined category of an unclear and imprecise positive category is even more unpromising as a basis for discussion. (In a book manuscript, I argue that in our discussion of self-consciousness, we should abandon talk of ‘non-conceptual contents of self-consciousness,’ and replace it with talk of a variety of more fine-grained types of contents, for instance, self-differentiations, self-images, self-schemata, self-concepts, self-ideas, self-conceptions, self-descriptions. I argue that the contrast between conceptual and non-conceptual is of little use, comparatively in an empirical assessment of the contents of self-consciousness.) One reason why Bermúdez rejects the argument for *the Symmetry Thesis* that takes its inspiration on Strawson and Evans’ focus on the generality constraint is because he thinks that the generality constraint is a constraint solely on concepts, and concepts are tied inextricably to language, so the only way to motivate an account of the Symmetry Thesis is by way of a criterion of non-conceptual contents. However, it will become transparent in the below, that I think the generality constraint is a constraint upon representations involved in self-ascription, apart from the consideration of whether those representations are conceptual or non-conceptual. So, the argument about *the Symmetry Thesis* is orthogonal to the consideration of the debates between conceptualists and non-conceptualists. Apart from my critique of Bermúdez’s paradox and non-conceptualism, Bermúdez does offer an *a posteriori* argument for the constitutive claim that is worth considering, to which I now turn.

Bermúdez argues, through an abductive argument, for “a constitutive link” between self-ascription and other-ascription. He argues that given that psychological self-awareness develops at approximately the same developmental stage during which there is a great deal of intersubjective interaction, then other-ascription “constitutes” self-ascription. I will argue in this section that Bermúdez’s *a posteriori* argument does not show that there is “a constitutive link” between self-ascription and other-ascription. There are two reasons for this. The first reason is that it is unclear what ‘constitutive’ is supposed to mean. The second reason is that a purely *a posteriori*—in this case, developmental— argument cannot show a constitutive link between the ability to self-ascribe and the ability to other-ascribe. A constitutive claim needs to be bolstered by a transcendental argument.

Bermúdez’s *a posteriori* argument appeals to what I discussed above as joint engagement:

- joint perception (infant and adult jointly attending to an object, e.g., a burning candle);
- joint action (infant and adult engaging in parallel actions, e.g., imitation in peek-a-boo); and,
- joint affection (infant and adult entertaining emotions that require intersubjective relations, e.g., pride and shame).

Bermúdez contends— via an abductive argument— the intersubjective relations that are realized in the joint phenomena constitute that self-ascription. In joint phenomena, an infant needs to be able to ascribe various psychological states to other subjects, for example, “x is seeing the bunny,” or “x is moving her hand,” or “x is scared.” Bermúdez

thinks that the possibility of joint phenomena shows that infants are capable of other-ascription of psychological predicates. His argument for *the Symmetry Thesis* is that infants develop the ability to self-ascribe psychological predicates around 9 months, and given that such joint phenomena occur at that time, then the best explanation for self-ascription is intersubjective interaction that is involved in the joint phenomena.

Although Bermúdez may be correct that these phenomena support the idea that pre-linguistic infants and non-linguistic animals possess primitive forms of self-consciousness, a mere appeal to these “joint phenomena” cannot provide an argument for *the Symmetry Thesis*. In order to make the argument for “a constitutive link” between self-ascription and other-ascription, one cannot merely appeal to the parallel and simultaneous development of self-ascription and other-ascription, but must show that the latter constitutes the former. (I will consider Bermúdez’s *a priori* argument in Chapter 6 and show that it is disconnected to the *a posteriori* argument that he presents.) The most a developmental argument could show is the falsity of a constitutive link between self-ascription and other-ascription.

Suppose we consider the claim “other-ascription is necessary for self-ascription.” Suppose that it were observed that self-ascription occurred in the absence of other-ascription. In this case, that constitutive claim would be falsified. However, a type of transcendental argument— not an abductive argument— is required to make the argument for “a constitutive link” between the ability to self-ascribe and the ability to other-ascribe, because experience only tells you what is the case, not what must be the case, as Kant argued. Constitutive claims are meant to support a necessary connection between two phenomena. My claim is that transcendental arguments serve this important

purpose in the philosophy of mind. They serve the purpose of showing that some ability is prior to another ability, and prior in a way that suggests that the latter ability constitutes the former.

The claim being made here is that shared mental states that are involved in joint perception, particularly the experience of persisting material objects, joint action, particularly the formation of we-intentions, and joint emotion, particularly the emotional response to potentially harmful objects or organisms, is dependent on a type of interpersonal awareness. Such interpersonal awareness of the other and the other's demands are the ground for constitution of the ascription of experiences in general. I will discuss the thesis in Chapter 6— the Persons Theory— according to which self-ascription and other-ascription develops in tandem and depends upon the recognition and acknowledgement of oneself as a person among persons. The joint experiences require the child to develop the ability for self-ascription and other-ascription of perceptions, actions and emotions— to develop the ability for both I-ascriptions, you-ascription and we-ascriptions. However, do these phenomena by themselves support the constitution claim?

Merely appealing to the phenomena of joint engagement cannot by itself support the constitution claim except by cooking the books in favor of self-other constitution. We can look at the normal course of infant development and always ask the open question, “But must we interpret it that way?” As I discussed in the introduction, there is the peculiar problem that there is quite weak data about isolated and feral children. However, it could be the case that these infants developed the capacity or ability to self-ascribe actions in isolation from others. As long as that is a rational and reasonable

question, and that question cannot be decisively falsified, then it remains a question that deserves an answer. But, how do we provide the answer? What we need to make clear are the assumptions that underpin the inference to the constitutive claim. To do so, let's talk about infants, monkeys and adult humans.

There is a familiar discussion in biology about whether ontogeny recapitulates phylogeny. The question is whether when we survey the development of an infant, in this case the development of self-consciousness, does this development provide a trajectory that enables us to account for the development of the species. When we look at the ontogeny of the infant's ability for self-ascription, we might find a similar account of the phylogeny of the human's ability for self-ascription (Bogdan (2000)). A similar question might be whether when we survey the development of an infant's capacity for self-ascription, does this developmental enable us to understand the ontology of self-ascription? Does the fact that certain types of interpersonal interaction occur prior to the development of self-ascription show that other-ascription constitutes self-ascription in general? If it did, then an infant could never develop self-ascription in the absence of others, because other-ascription would be a necessary, meaning, a non-negligible condition. The claim, therefore, is that ontology does not recapitulate ontogeny.

It might be objected that my talk of self-ascription in general focuses merely on the semantics or conceptual meaning of 'self-ascription.' But, by asking about self-ascription in general, I am not merely considering the semantics or conceptual meaning of 'self-ascription.' Instead, I am asking about the theoretical assumptions that we make about the concept of self-ascription when we attempt to operationalize that concept. Any inquiry into a phenomenon needs to reflect on the grounds of its operationalization. I am

not merely considering terms, labels or concepts, but instead engaging in substantive theoretical considerations about how those terms, labels or concepts are legitimated by the phenomenology. But, what exactly does the constitution claim amount to then?

The claim is that the ability to ascribe experiences to other subjects like oneself is necessary for the ability to ascribe experiences to oneself. What does ‘necessary for’ mean in this claim? Let’s take a more familiar example in order to make this locution more clear and precise. For example, moving one’s arm is necessary for throwing a baseball. It is thus a necessary condition of throwing a baseball that one move one’s arm. It is this sense of ‘necessary for’ that I have in mind. X is necessary for Y_{def} : Y cannot obtain without X . Substituting the constitution claim, the claim is that other-ascription is necessary for self-ascription, namely self-ascription cannot obtain without other-ascription. Throughout the dissertation, I will use the terms, ‘necessary for,’ ‘requires,’ ‘necessary condition of’ interchangeably. It is this analysis of these terms that should be kept in mind.

I should say that the types of evidence that figures like Bermúdez present definitely support the constitution claim, but I am arguing that the evidence cannot be the sole basis for it. However, I should pause here to say a bit more about the types of phenomena that are central for our understanding of joint experiences. It is common to our experience that caregivers and infants have a certain mutual experience. We jointly attend to objects in visual perception. We act in accordance with each other. We react similarly to objects in our environments. What are relevant for our discussion, however, are those cases in which an infant is other-ascribing experiences. It will become clear

that the joint phenomena that I am focusing on are all we-phenomena, in the sense that they involve contents that have a ‘we’ in the subject place.

For instance, when a caregiver and infant are looking at a spider, rocking in a chair, or reacting to a knife, then the infant might have the experiences “we are seeing a spider”, “we are rocking in a chair”, and “we are reacting to a knife”, respectively. The tendency is to think that joint, shared, or mutual experiences are the complex of two fully-formed subjective experiences. But, there is little reason to be committed to this assumption. If the constitution claim is correct, then the claim that “the I is a We; the We, an I” should not commit the infant to possessing a fully-formed subjectivity. As I will discuss in Chapter 6, this interpretation of the joint phenomena is too robust, because a more minimal interpretation better captures the phenomena. I now turn to the final consideration for why the developmental phenomena are not sufficient to substantiate the constitution claim.

Why do I say that a developmental argument cannot do the work that Bermúdez wants it to? The answer to this depends upon a proper articulation of what the import of the constitution claim is supposed to be. It is not sufficient to defend the constitution claim by suggesting that interpersonal interaction precedes the acquisition of self-ascription. To draw the inference to the constitution claim from the mere correlation of these two sets of phenomena— the interpersonal interaction, on the one hand, and the ability to self-ascribe, on the other— would be to commit a version of the *post hoc* fallacy. The *post hoc* fallacy is a fallacy of reasoning that because Y occurs after X, that therefore, Y is the cause of X. But, it is widely accepted that correlation is not causation, despite Hume’s constant conjunction account of causation. Bermúdez, however, does not

merely claim that other-ascription (Y) is the cause of self-ascription (X), but that other-ascription (Y) constitutes self-ascription (X). What needs to be shown in addition to temporal precedence/subsequence is that Y is prior to X in a deeper sense. The slogan for this point might be: causality *qua* correlation is not sufficient for constitution. I will argue that in order to make the constitution claim, we have to legitimate what I will call a ‘no priority symmetry claim’ about the relation between self-ascription and other-ascription.

As I discussed in the introduction, my account in the dissertation provides a genuine alternative to the two *asymmetrical* accounts of the self-other relation: the Simulation Theory and the Theory Theory. The Simulation Theory privileges the self; and the Theory Theory privileges the other. But, each does so at the expense of the other pole. In order to make this argument, however, I have to accomplish two aims. The first aim is to show that the developmental argument is insufficient to establish the constitution claim. The second aim is to show that the constructive argument is necessary to establish the constitution claim. I will argue that in order to conclude that the ability to ascribe experiences is comprised interpersonally, both of these aims need to be met. I turn to meeting the first aim by backing up to consider how the constitution claim is an answer to a how-possible question about self-ascription.

2. How-Possible Questions about Self-Ascription

When Bermúdez considers this form of self-consciousness— although he does not stress the point— he asks a how-possible question about self-ascription. The how-possible question is “How is self-ascription possible?” Both P. F. Strawson’s and Evans’ accounts of self-ascription are inspired by this sort of question, because they both take it

to be philosophically significant how self-consciousness arises. (One recent discussion of how possible questions is in Cassam's (2007) The Possibility of Knowledge, in which he discusses a framework for answering how-possible questions that inspires the account provided here.) They both think that a constructive argument is required to link self and other. In this sense, they both agree that the constitution claim requires a transcendental argument. In this section, I will discuss the context for the origin of the how-possible question, as this will provide a way to meet my first aim, namely to show that an *a posteriori* argument is insufficient to meet make the constitution claim. The how-possible question arises in the context of skepticism about our knowledge of other minds. I will briefly discuss the basis for this skepticism. Then, I will provide an answer to the how-possible question about self-ascription.

The how-possible question about self-ascription considers what makes possible self-ascription is I-thoughts. This brings us to the question, "How are I-thoughts possible?" I consider four responses: (1) *the non-answer* that I-thoughts are innate; (2) *the means answer* that I-thoughts develop through interpersonal interaction; (3) *the obstacle-overcoming answer* that addresses difficulties in accounting for I-thoughts and (4) *the constructive answer* that I-thoughts are constituted by we-thoughts. I provide objections to the first three answers and suggest that the last answer requires a transcendental argument for *the Symmetry Thesis*. But, before we consider these answers, I need to provide some background about how self-consciousness relates to the problem of other minds, because it is in the context of that problem that awareness of other minds seems difficult.

The problem of other minds arises for views of self-consciousness that assume that we have direct and immediate knowledge of our own minds and indirect and mediated knowledge of other minds. For example, if I think “I have a headache” then my self-ascription of that mental state is know directly and immediately in my case. But, if someone else has a headache, then I can only say “You have a headache” or “She has a headache” through an indirect method, namely I can only do so through inferences from the way my experience is to how yours or hers might be. The typical indirect method is via analogical inference. For example, when I have a headache, then I engage in verbal and non-verbal behaviors that correlate with headaches. When I see you or her engaging in those same (or relevantly similar) behaviors, then I make an analogical inference that you or she must also have a headache. But, this inference is only problematically inferred, for there are all sorts of defeaters that might arise. You or she might not have a headache. You or she might be faking it. Or in the extreme case, you or she might be a robot or a zombie, for all I know. In this case, then the analogical approach to the problem of other minds assumes that the content of “I have a headache” has meaning in solely first personal terms.

However, I will be considering neither the problem of self-knowledge nor the problem of knowledge of other minds. As I discussed in the introduction, I take the problem of other minds to be based on a faulty conceptual assumption. The faulty assumption is that the content of experience is intelligible as a content that is solely first personal. This mistake has been diagnosed by Anita Avramides (2001) in her account of the problem of other minds. She argues that the epistemological problem of other minds presupposes the conceptual problem of other minds. *The Epistemological Problem* is a

problem of how we are justified in believing that others have minds or that others have experiences like ours. *The Conceptual Problem* is a problem of how we possess the concepts of other minds or the concepts of other individuals' mental states. As I will discuss in the conclusion, once the argument for *the Symmetry Thesis* can be made good, then the problem of knowledge of other minds should not arise. As Strawson the Elder has argued (1985), the transcendental argument for *the Symmetry Thesis* enables us to ignore or dissolve the problem of other minds rather than providing us with a direct solution.

However, the how-possible question also arises in another context. It also arises because it can seem to be a mystery how self-consciousness could arise as a product of the natural world. We are considering self-consciousness to be self-ascription of experiences, and self-ascription of experiences is understood in terms of I-thoughts. However, I-thoughts seem to have a peculiar type of reflexivity that is not found in the world otherwise. Thinking about a rock is thinking about an object; but, thinking about oneself is thinking about both a subject and an object. For example, Hofstadter (2007) has called this peculiar type of reflexivity 'the strange loop.' In this case, we have then two obstacles to making sense of how self-ascription is possible. That is why the question is raised, "What makes self-ascription possible?"

It will be helpful to consider a recent treatment of how possible questions by Quassim Cassam (2007: Chapter 2). Cassam elucidates three types of answers to how-possible questions. The first answer to a how-possible question is *the means answer*, which presupposes that the question is asking for the means to achieve something or other. If one asks, "How is it possible to get from Rapid City to Devils Tower?" one

might answer by providing the means to doing so, namely “Take the Interstate 90 west crossing into Wyoming; then, take US 14 north; turn left onto US 24 north; drive 6 miles and Devils Tower is on your left; you can’t miss it.” The second answer to a how-possible question is *the obstacle-overcoming answer*, which presupposes that the question is asking for an answer about how some obstacle or other is overcome. If one asks, “How is it possible to get from Rapid City to Devils Tower?” they might be asking this question because there is a wildfire consuming Blackhawk, SD (the first town west of Rapid City on the Interstate 90). In this case, the answer also provides means, but the answer, in addition, provides means that offer a way to overcome an obstacle, namely it is not by way of the Interstate 90, but by way of a detour that will allow you to arrive at Devils Tower by another way. The third answer to a how-possible question is the offering of an argument for a necessary condition, which I will call *the constructive answer*. If one asks, “How is it possible to get from Rapid City to Devils Tower?” in the context of an assumption that they not leave Rapid City, then the third answer is called for. Namely, you have to respond (after the requisite stare of bafflement) that it is a necessary condition that one leave Rapid City in order to get from Rapid City to anywhere else. Notice that the third answer via a necessary condition also provides a means and overcomes an obstacle, but it provides a deeper answer because it makes intelligible the types of answers that are available and it limits the types of questions that are possible.

When I ask, “What makes possible self-ascription?” what type of question-answer pair am I concerned with? While I will discuss the first and second to some degree, it is really the third question-answer pair— the how-possible/necessary condition pair— that

is my central concern. As I discussed above, I-thoughts make self-ascription possible, so we are concerned with answering the question, “What makes I-thoughts possible?” I will consider construing this question as requiring a means answer and an obstacle-overcoming answer, and then discuss why the question requires a constructive answer. But, before I do, I need to take into account what I take to be a non-answer.

Occasionally, when philosophers discuss self-consciousness, they tend to respond that self-consciousness is probably innate. In this case, the claim would be that I-thoughts are innate. However, this is nothing but a non-answer, since if I-thoughts were innate, then infants would be capable of self-ascription from birth or even prior to birth. But, this is not the case, since self-consciousness is a mental phenomenon that develops. Also, the ability for I-thoughts cannot be innate because infants that do not have normal development have impairments in I-thoughts. So, to this extent, I agree with Bermúdez’s constraint on attributing contents to an organism (1998), that those states or processes are something that can be acquired in the normal course of development, which he calls ‘the acquisition constraint,’ according to which “if a given cognitive capacity is psychologically real, then there must be an explanation of how it is possible for an individual in the normal course of human development to acquire that cognitive capacity” (1998, p. 19).

Let’s consider the means answer. The means answer suggests that there is some means or way that I-thoughts develop that explains what constitutes them. I will consider several means by which self-ascription via I-thoughts develops. The first condition is that the creature has a sensibility (Strawson 1959, p. 84–88). A creature uses its sensibility as a means in the sense that it receives input from the world, which determines

its perception, action, emotion and thought. The second condition is that the creature locates itself in space and time (Strawson 1959, p. 84–88). A creature uses self-location as a means in the sense that it could not determine what it was ascribing experiences to unless it could locate itself in space and time. The third condition is that the creature be able to orient itself by tracing out a line in space and time (Strawson 1959, p. 84–88). A creature uses self-orientation as a means in the sense that it has a life trajectory. Without a life trajectory, the creature would not be able to situate itself in a form of life. The fourth condition is that it be capable of distinguishing between appearance and reality (Russell (1995, p. 134)). The fifth condition is that it know-which embodied object it is (Cassam (1994, p. 108)). Each of these conditions is used as means to determining the empirical grounds or basis of self-ascription. Why aren't these empirical conditions sufficient for determining self-ascription? I will argue in Chapter 6 that the creature needs to recognize and acknowledge a set of norms that govern perception, action, emotion and thought that supersedes these empirical conditions. While the empirical conditions are necessary for the recognition and acknowledgement of oneself as a person among persons, they are not sufficient, because in order for creature to be receptive to norms, they need to interact with other persons. So, while these empirical conditions are necessary, they are not sufficient. It is in this sense that a means answer is not sufficient to answer the how-possible question.

Next, let's consider *the obstacle-overcoming answer*. The obstacle-overcoming answer suggests that there is some obstacle to the development of self-ascription by way of the other-ascription of experiences. The assumption is that there is a *prima facie* obstacle to the description of self-ascription as dependent upon other-ascription, because

of the problem of other minds. In particular there is the obstacle that I am not aware of other minds in a direct and immediate way. Like the wildfire in Blackhawk is an obstacle to the route to Devils Tower, the problem of other minds presents an obstacle to understanding self-ascription as dependent upon awareness of other minds. But, the problem with the obstacle over-coming answer is that it is tailored to the obstacle that is presented. So, presumably, there will always be another obstacle waiting around the corner. For instance, the skeptic about other minds can always claim that there are other defeaters to the links between self-ascription and other-ascription. So, it cannot be an obstacle-overcoming answer that is required. Instead, it is a constructive answer that is required.

The constructive answer provides a transcendental argument for the constitution claim to the effect that Y is required for X. The condition is necessary but it is necessary in a way that is constitutive of the phenomena in question. There is as yet no clear articulation of an argument for this sort of constructive answer to the how-possible question about self-consciousness. As I said above, Bermúdez hints that there are resources for such an argument in the work of P. F. Strawson and Gareth Evans, but his skepticism of that strategy led him to put that argument aside. (One exception is Sebastian Rodl's (2007) account in Self-Consciousness, however, the concept of the constitutive is never made clear and precise in that text, but the notion of constitutivity is merely assumed to be transparent.) So, in order for a creature to be capable of self-ascription of perceptual, agentive or emotional experiences, it is required that the creature be capable of ascription of perceptual, agentive or emotional experiences to others. What makes self-ascription the kind of phenomenon it is being tied to other-ascription in the

way that it is? This is what it means for it to be constitutive. However, if the constitutive answer requires a transcendental argument, then I need to defend the use of transcendental arguments against typical objections.

3. Transcendental Arguments

I will now focus on the issue of transcendental arguments, since there have been objections to the use of transcendental arguments in philosophy in general. It might be objected that transcendental arguments are fatally flawed because they commit the transcendental arguer to a kind of philosophical hubris. Transcendental arguments are generally defined as partly non-empirical anti-skeptical arguments that are concerned with arguing for the necessary conditions of some type of knowledge, language or thought. In Kant's terminology, transcendental arguments answer the question of what makes experience in general possible (or any especially important aspect of experience possible), by arguing for what is necessary for that experience in general (or any especially important aspect of that experience). As I discussed in section 2 above, the constitutive answer answers the question "What makes possible I-thoughts?" with the answer that the ability to entertain we-thoughts are necessary for the ability to entertain I-thoughts. It should be clear, however, that this claim need not be strictly non-empirical and need not be concerned with anti-skeptical aims. The other option is to provide a transcendental argument.

There has been a lot of interesting recent debate about the use of transcendental arguments in general, but little discussion about the use of such arguments in philosophy of mind and the philosophy of self-consciousness. Austin (1939) is the first to use the term, as far as I can tell. It is in the context of a discussion of *a priori* understanding.

Taylor (1975) defines transcendental arguments as “arguments that start from some putatively undeniable facet of our experience in order to conclude that our experience must have certain features or be of a certain type, for otherwise this undeniable facet could not be” (1975, p. 151). Barry Stroud (1968; 1999) provides a skeptical challenge to the use of transcendental arguments, which I will face below. Rosenberg (1975) revisits the notion of a transcendental argument given Stroud’s criticisms with particular emphasis on Kant’s use of transcendental arguments. Cassam (1987) provides an account of transcendental arguments in light of Kant’s notion of transcendental synthesis. In Cassam’s (1997) account of the dependence of self-consciousness upon the consciousness of the body, he suggests that modest transcendental arguments are all he intends to use. Anthony Brueckner (1996) argues that even modest transcendental arguments fail. Paul W. Franks (2005) provides an analysis of the role of Post-Kantian transcendental arguments. I will not summarize each figure’s account of transcendental arguments, but instead, will discuss how transcendental arguments play a role in the dissertation, making sure along the way to register agreements and disagreements with central figures in the debate about the merit of such arguments. Let’s begin by considering the logic of a transcendental argument.

In the history of philosophy, Kant provides the most famous transcendental arguments, although he never uses the term. Kant (1781, §§ A19/B33-A49/B73) argues in the *Transcendental Aesthetic* for a type of transcendental empiricism. The major claim of transcendental empiricism is that space and time are *synthetic a priori* conditions of perceptual experience. The structures of space and time are the non-negligible conditions of perceptual experience in general. So, any creature that we might consider— human,

non-human, robot or alien— if that creature is to count as having perceptual experience, then that creature must have perceptual representations that possess spatiotemporal content. Kant argues in the transcendental deduction that the pure concepts or categories are the necessary conditions of thought about and cognition of nature. So, any creature that we might consider— human, non-human, robot or alien— if it is to count as making judgments about the world— in thought or in cognition (meaning thought that entails an a material object) — then that creature must have categorial representations that structure its judgments. (Kant also argues in the analogies that we need to assume substance metaphysics, the reality of causation, and the principle of simultaneity. I will not rehearse these arguments here.) However, I will consider in the beginning of Chapter 2 one pivotal argument in the B deduction— the argument for requirement of the synthetic unity of apperception for intentionality— which I take to be a transcendental argument. What might be some objections to this type of argument? I will consider two objections that I take to be unrelated to my project and make sure to put them aside.

The first unrelated objection is what might be called ‘the modality objection.’ (This objection was put to me by David Pereplyotchik in a variety of forms.) The modality objection to transcendental arguments is the most common objection to transcendental arguments. Since transcendental arguments for constitution claims (as I have defined them) rely upon a claim about necessary conditions, it might be objected that this commits me to an implausible type of metaphysical necessity and possibility. One might argue that if there are no metaphysical necessities, transcendental arguments cannot be valid. However, the transcendental arguer need not be committed to the existence of modal facts. To say that Y is necessary for X is not to claim that something

exists necessarily (or even possibly); ‘Y is necessary for X’ is different than ‘Y is a necessary being.’ Let’s consider a naturalistic example to make this transparent.

There are occasions in the physical sciences in which natural laws are couched in similar terms. For instance, consider the natural law that if water is to boil at sea level, then it must be heated to 210 degrees Fahrenheit. In this case, the type of ‘Y is necessary for X’ that is involved does not require any metaphysical necessity, but instead merely the type of necessity that relates boiling and boiling point. This is the type of necessity that is implied by the constitutive answer. Another more theory-laden way to make this point is to say that modal skepticism is compatible with acceptance of the import of necessary conditions, since the commitment to necessary conditions only relates phenomena, but does not commit one to the existence of metaphysical necessities or possibilities.

The second unrelated objection is what might be called the conceptual scheme objection. The conceptual scheme objection is an objection to a particular type of construal of transcendental arguments under which the purpose of transcendental arguments is to argue for analytic truths. Both P. F. Strawson (1966) and Jonathan Bennett (1966) subscribe to the idea that transcendental arguments make analytic connections between concepts. Stephan Korner (1974) has argued that if transcendental arguments explicate analytic connections between ideas in a conceptual scheme, then there is no way to ensure that that conceptual scheme is the only one. So, while the conclusion that self-ascription and other-ascription are constitutively dependent on each other is true in our conceptual scheme, it may not be in the conceptual scheme in a non-Western culture, for instance in the Hindu notion of the “Aham Brahmasmi” or “I am the

Absolute Truth,” there is no world external to the self or I. So, connections might be unique to the non-Hindu conceptual scheme, but might not hold in other conceptual schemes, for instance in the Hindu conceptual scheme. However, I do not accept the claim that transcendental arguments demonstrative analytic truths, since I agree with Quine (1953) that there are no strictly conceptual truths. Instead, transcendental arguments demonstrate synthetic truths, because they are constructive arguments that legitimate a connection between an experiential premise and a transcendental conclusion. The methodology involves presenting a synthetic constructive argument.

However, these are not the only objections and there are objections that are relevant to my approach. Some have argued that transcendental arguments imply implausible positions with respect to our understanding of the phenomena they attempt to elucidate. These objections are worth considering more closely, because the response that is required is not merely one of sloughing off metaphysical commitments, as was done above in the rejection of metaphysical necessity and analyticity. Barry Stroud (1968) has argued that transcendental arguments are invalid because they imply either idealism or verificationism. One quick response to Stroud is that simply because the use of an argument entails an implausible ‘-ism,’ that is not a decisive objection to the argument itself, partly because any ‘-ism’ is true to the phenomena under some interpretation. For example, everyone is a solipsist under some interpretations. However, let’s take a look at Stroud’s argument in more detail. To do so, we consider the structure of a transcendental argument:

- (1) A claim P about experience, the truth of which is accepted.

(2) The truth of a non-experiential proposition Q is a necessary condition of P.

(3) Therefore, Q.

Remember that I discussed above the idea that the constitution claim plays a role in a larger argument concerning the problem of other minds. Let's substitute the content of the problem of other minds in this transcendental argument structure. Premise 1 above is an experiential premise. Premise 2 above expresses a necessary condition. Premise 3 above is the transcendental conclusion.

(EP) We are aware of our own minds (the experiential premise).

(NC) Being aware of other minds is a necessary condition of being aware of our own minds (the necessary condition).

(TC) Therefore we must be aware of other minds (the transcendental conclusion).

Stroud argued that claims like EP may only appear to be true, and therefore, that claims like TC do not follow. At best, transcendental arguments show us what we must believe about appearances, rather than showing us the layout of nature. But, I take the existence of self-consciousness to be a Moorean fact, namely that self-consciousness does not require an existence proof. But, this does nothing to establish the relationship between awareness of oneself and awareness of other minds. The proponent of transcendental arguments then has to show how we connect what we believe about appearances to the way that things are in reality. Stroud argues that this can only be done if the proponent accepts either idealism or verificationism. I consider each of these in turn. If Stroud is

correct in his critique (which I think he is), then I will have to accept a grazing of one of the bullhorns of Barry's dilemma.

In the former case, the idealist suggests that the way that we talk or think about appearances determines the nature of reality. So, the claim would be that the way we talk and think about self-ascription as such would determine how we understand ascription of experiential predicates to ourselves. In this case, the contrast is between idealism about other minds and realism about other minds. The idealist about other minds would be committed to the view that since we talk, think and believe that there are other minds, there are other minds. In this case, our ways of thinking would determine the nature of the reality that we inhabit. The realist about other minds would be committed to the view that there are other minds independent of our talk, thought and belief about them, that they exist in an evidence-transcendent way.

But, I will not be grazed by this horn. I will argue that the constitution claim commits us to the reality of interaction with other minds. The reason for this is the idealism about other minds is really no more successful than is the argument from analogy. I agree with Stroud that the objection from idealism is something to be avoided. For what import could possibly arise from the idea that we think there are other minds, even if what we think is constitutive of reality. It is of no great import that in order to have the ability for self-ascription we have to talk, think or believe that there are other creatures like ourselves that have or possess experiences like ours. To achieve only that result would be worse than a consolation prize. I think real interpersonal interaction is a necessary condition for our ability and capacity to self-ascribe experiences, and I will assume that as a target for the constructive argument for *the Symmetry Thesis*. One way

to capture this is with the following claim: the factive perception of other persons and other persons' mental states is a necessary condition for self-consciousness. We possess a biological nature as creatures that have sensibilities and meet the empirical conditions outlined above. Upon interacting with other persons in a direct and unmediated way, we are forever altered and become able to do things that we wouldn't and couldn't do otherwise. In this sense, I am committed to the reality of other minds, so I do not wish to accept idealism about self-consciousness, about the necessary condition between self-ascription and other-ascription, nor about awareness of other minds. What about the other horn of Barry's bull?

The other horn of Barry's bull is the verificationist objection that suggests that the meaning of our terms that we apply to appearances can only be secured given there are experiential criteria that determine their application. As I discussed above, if the question is "What makes self-ascription possible?" then the immediate answer is that "I-thoughts make self-ascription possible." The question is then, "What makes I-thoughts possible?" The verificationist objection suggests that the sense and meaning of I-thoughts are determined by experiential criteria. The reason that the verificationism objection in general is devastating is because: (1) we cannot specify the conditions that verify the principle of verification itself; (2) many propositions are meaningful even though we cannot specify their methods of verification.

It is this horn of the dilemma that will graze me, because I agree in general that the verificationist critique is devastating for the reasons outlined above, but that the account of verificationism that I accept about I-thoughts is not problematic. So, wayward bull rider that I am, I will survive to tell the tale. In order to deal with this objection, I

will have to accept a form of verificationism, but one of an unproblematic kind. First, the general problem with the verification principle is no problem at all for the account of I-thoughts offered here. By being committed to a limited verificationism about the I-think, I am not committed to verification about the significance of all terms or concepts.

Second, while there may be terms or concepts for which we cannot specify their method of verification, this is not so given the I-think. To say that the I-think is dependent upon a method of verification is to say that there are criteria for the application of that representation. P. F. Strawson discusses the principle of significance in The Bounds of Sense: “There can be no legitimate, or even meaningful, employment of ideas or concepts which does not relate them to empirical or experiential conditions of their application” (1966, p. 16).

Similarly, in Evans’ the Varieties of Experience, Evans (1982, p. 93–100) accepts that his view is a kind of verificationism, to the extent that there is a two-tiered account of thoughts. The non-basic level is the level of the sense of thought. The basic level is the level of the demonstrative identification of the content of the thought. This non-basic level is understood in terms of *the fundamental ground of difference* between the object of the thought and other objects, and this ground is in empirical and experiential conditions. In this respect, the account that I will offer here will be modeled on the above accounts. To say that the “I-think” depends upon empirical conditions is to accept an unproblematic form of verificationism, since it is unharmed by the bull’s horns: 1) this form of verificationism about ‘I-thoughts’ doesn’t commit me to a general verificationism, in which case the principle of verificationism does not require a verification condition; 2) this form of verificationism about ‘I-thoughts’ doesn’t commit

me claiming that propositions are meaningful or not meaningful, because the contribution of ‘I-thoughts’ to mental economy is a functional role, not a role that is determinative of the content of propositions. It will become clear that one of the empirical conditions of the I-think is the criteriological connection that it has to the we-think, however this will need to be argued for.

In response to these two objections above, Stroud (1999) has proposed that we be satisfied with more modest transcendental arguments. Modest transcendental arguments attempt to show the indispensability of some belief, thought or concept. Such arguments have been called “belief-directed” transcendental arguments, rather than “world-directed” transcendental arguments, because they only attempt to show that we should believe something to obviate incoherence or inconsistency. Given that I am willing to accept some form of verificationism about the I-think, it is clear that I think that Stroud’s modesty is overrated in this case. I do think that there are occasions where modest transcendental arguments are called for, for instance, in arguments for the dependence of thought upon perception, but I will not discuss these arguments here. The arguments that I will employ in the dissertation should be considered “world-directed” transcendental arguments. I am arguing for a substantive claim in the philosophy of mind, not merely about how we should talk, think, or what we should believe. The arguments for *the Symmetry Thesis* are meant to instruct us about how we should talk, think or what we should believe given our assumptions, but they are meant to illuminate the primitive grounds or basis of self-ascription as well. That is to say that inquiry into the nature of cognitive abilities is equally inquiry into the way the world is. In this sense, while they

are theoretical arguments that are belief-directed, importantly, they are also empirical arguments that are world-directed.

Now that I have defended the use of transcendental arguments, what is the argument for *the Symmetry Thesis*? My purpose is to provide a thorough defense of the Symmetry Thesis through the following argument. As I mentioned above, I stipulate the kind of self-consciousness that I am concerned with.

- 1) Self-consciousness is the self-ascription of experiential predicates and the self-ascription of experiential predicates has a subject-predicate structure.
- 2) I argue that because of its structure, self-ascription has to meet the generality constraint.
- 3) If self-ascription requires meeting the generality constraint, then this requires the ability to generalize the subject and the ability to generalize the predicate involved in self-ascription.
- 4) Putting aside generalizing the predicate, generalizing the subject of self-ascription depends crucially on identification of the subject of experiential predicates, namely self-ascription depends upon self-identification.
- 5) Self-identification cannot be done in terms of experiential predicates alone, but instead in terms of the subject of those predicates.
- 6) Thus, a different principle of self-ascription must be used— what I call the “the Persons Principle”: in order to be able self-ascribe experiences, one must identify the subject of self-ascription as a person among persons.

7) Creatures need to have the ability for the recognition and acknowledgement of oneself as a person among persons in order to generalize the subject.

C) Therefore, the ability to self-ascribe requires the ability to other-ascribe, i.e., the ability to ascribe experiential predicates to other subjects.

Thus, the Symmetry Thesis follows.

In the next chapter, I begin the constructive argument for *the Symmetry Thesis*. *The Symmetry Thesis* is the claim that the ability to self-ascribe experiential predicates requires the ability to ascribe experiential predicates to other individuals like oneself. The first premise of this argument makes a claim about the nature of self-ascription: (1) *self-ascription is the ascription of experiential predicates in the subject-predicate form*. The first purpose of the next chapter is to defend premise one by critiquing accounts of self-ascription that do not hold that view. The second purpose of the next chapter is to argue that if self-ascription has a subject-predicate structure then it must meet the generality constraint, which is the second premise of the argument for *the Symmetry Thesis*. I present an argument that self-ascription must meet the Generality Constraint.

Chapter 2

“A subject who can only have indexical thoughts about herself will obviously not have a full understanding of "I". She would not be able to entertain context-independent thoughts about herself or think about herself as an instance of a category of beings similar to herself. These points are emphasized by Evans when he asserts that thoughts about oneself must conform to the Generality Constraint.”
— Ingar Brink (1999)

I See a Sheep: **Self-Ascription, the Subject-Predicate Structure and Generality**

In the previous chapter, I defended the use of transcendental arguments in the inquiry into the nature of self-consciousness. I argued that in order to substantiate a constitutive claim about self-consciousness— the ability for self-ascription is constituted by the ability for other-ascription— one needs to argue that the latter is necessary for the former. ‘Necessary for,’ in this case, is understood as arguing for a transcendental condition, which requires a constructive argument. While correlation and causal priority, of course, bolster the strength of a transcendental argument, one cannot defend a constitutive claim unless one argues for a necessary condition. I responded to various objections to the use of transcendental arguments in the philosophy of self-consciousness, specifically, the objection that this leads to either idealism or verificationism, both of which are unbearable implications. I argued that ‘I-thoughts’ do depend upon empirical criteria, and so that in this sense, verificationism is the horn of the dilemma that I am grazed by.

In this chapter, I begin my constructive argument for *the Symmetry Thesis*. *The Symmetry Thesis* is the claim that the ability to self-ascribe experiential predicates requires the ability to ascribe experiential predicates to other individuals like oneself. The first premise of the argument makes a claim about the nature of self-ascription: *(1) self-ascription is the ascription of experiential predicates in the subject-predicate form*. Strawson (1959; 1966), Evans (1982), and Bermúdez (1998), all accept this account of self-ascription as an assumption, so my contention will not be with them. However, these thinkers do not provide the only accounts of self-ascription that hold this assumption. In addition, there are other accounts of self-ascription that do not hold the assumption that self-ascription is the ascription of experiential predicates in the subject-predicate form.

Thus, the first purpose of this chapter is to defend premise one by both showing how prevalent this assumption is in accounts of self-ascription and otherwise critiquing accounts of self-ascription that do not hold that view. I discuss other views of self-ascription that support the premise and consider the opposing accounts. The chief argument against the assumption that self-ascription has a subject-predicate structure comes from Humean eliminativists and Neo-Expressivists about self-ascription. The second purpose of this chapter is to argue that if self-ascription has a subject-predicate structure then it must meet the generality constraint, which is the second premise of the argument for *the Symmetry Thesis*. In order to defend this second premise, I consider arguments that support this premise by focusing on the content of the predicates. The focus of these arguments is that there is something about particular experiential predicates that require generality. I discuss two theses about self-ascription support the

idea that the generality constraint must hold: (IPT) the interpersonal predicates thesis and (CST) the constancy of sense thesis.

However, I argue that these theses only *presuppose* the generality constraint, and cannot be used to argue for it. I pause to suggest that there might be a shift of focus merited about self-ascription. On the account of *the Symmetry Thesis* that I recommend, the role of the identification of the subject of self-ascription is what is significant, not the content of the predicates of self-ascription. The claim is that *the Symmetry Thesis* is better defended through an argument from structure, rather than an argument from content. To offer an example to clarify, when one thinks “I am seeing a sheep,” it is not the content of the experiential predicate “...seeing a sheep” that is the focus on the proper argument for *the Symmetry Thesis*, but instead what is presupposed by the I-thought. What I will focus on is the requirement of generalizing the subject, not generalizing the predicate of self-ascription. I present an argument that self-ascription must meet the Generality Constraint. I then consider objections to the argument, and reply that they do not overturn the generality constraint argument. I, therefore, conclude that self-ascription must meet the Generality Constraint.

1. Models of Self-Ascription, Hume’s Error and Neo-Expressivism

As I outlined in the introduction, the focus of the dissertation is a particular kind of self-consciousness: the self-ascription of experiential predicates with the subject-predicate structure. An example of a self-ascription is thinking, “I am seeing a sheep.” The experiential predicate is “...seeing a sheep.” The subject is “the I” of the self-ascription. In the argument that I outlined in the introduction, the first premise is that self-ascription has a subject-predicate structure. While not all accounts of self-ascription

accept this, there are those that at least do not openly oppose it. While I will not take up these accounts in detail (for that would take me too far afield), I will mention the alternative accounts to ameliorate any concern that those accounts are being ignored. I briefly suggest in each case why each account admits that self-ascription has a subject-predicate structure.

I have said that the best example of self-ascription is “I am seeing a sheep.” When one thinks, “I am seeing a sheep,” one is ascribing to oneself the visual experience of seeing a sheep. One is not merely seeing in this case. In addition, the seeing is a state or process that is for one. If I think “I am seeing a sheep,” the predicate is the content of the experience of seeing a sheep and the subject is the individual that “the I” denotes in thought. The predicate has the content of visually experiencing a sheep. The subject has the content of that to which the experience is being ascribed. The structure of self-ascription is the subject-predicate structure, because it would not count as a self-ascription unless there were a distinction between the ascriptive predicate, on the one hand, and the subject to which that ascription is ascribed, on the other.

In order for an individual to self-ascribe an experiential predicate, that individual need not say or speak the words, “I am seeing a sheep.” Not all self-ascription is linguistic self-ascription. As I argued in the chapter on methodology, if self-ascription were always linguistic, then we could not self-ascribe unless we linguistically self-ascribed. However, this is clearly not the case when we have self-ascriptions in thought alone without expressing them. Everyone is familiar with thinking, “I am sad,” for instance, and not expressing that sadness in overt speech. But, this should not be taken to rule out the claim that self-ascriptions have a sentential structure. Neither is it required

that one introspect that one is seeing a sheep, and introspectively self-ascribe “I am seeing a sheep.” Not all self-ascription is introspective self-ascription. As I discussed in the introduction, for this reason, the account of self-ascription on offer here is not dependent upon introspection. Also, in order for an individual to self-ascribe an experiential predicate, that individual need not consciously self-ascribe that one is seeing a sheep. Not all self-ascription is conscious self-ascription. On the account of self-ascription offered here, therefore, self-ascription is a type of thinking about one’s experiences that is not necessarily expressed linguistically, is not necessarily introspected and is not necessarily conscious.

But, what makes this account of self-ascription an account of self-consciousness? And what argument and evidence is there that such a form of self-ascription exists? Because one is ascribing one’s own experience to oneself, then one is conscious of oneself, since being self-conscious is usually understood as being aware of oneself as oneself. In this case, being aware of oneself as oneself is being aware of oneself as the subject entertaining the experience of seeing a sheep, being the center or locus of one’s experiences. Most philosophers and psychologists mean by ‘self-consciousness’ something like the ascription of one’s experiences to oneself. For example, one is self-conscious if one is ascribing to oneself an experience by thinking, “I am hungry.” In this case, one is self-ascribing the experience of being hungry by thinking I-thoughts about one’s experiences. Such I-thoughts have a particular functional role in one’s thought in general.

It is usually assumed that I-thoughts have import for perception, action and emotion in ways that other thoughts do not, as Perry (1979) effectively argued. I extend

this account in the following. I-thoughts have import for perception in being related to here-thoughts and now-thoughts, which illustrates their basis in empirical criteria (Evans (1982: Chapter 7)). For example, if I think “I am seeing a sheep,” then my self-ascription is dependent upon ways of gaining information about the objects that are present in the here and now, for instance through the spatiotemporal structures of perception. I articulate the kinds of empirical conditions that are placed on self-ascription below. Also, my ways or modes of gaining information through perception are dependent significantly on my I-thoughts, for awareness of my spatiotemporal position is grounded in my capacity to have I-thoughts. For example, unless I were able to think “I am seeing a sheep,” I could not think “this sheep is here and now.” As such, egocentric and allocentric representation in perception are two sides of the same coin (Grush (2000, p. 90)).

I-thoughts also have immediate import for action. For example, if I think “I am going to be eaten by a bear,” then this has immediate import for action in a way that “James is going to be eaten by a bear” may not, because I may not know that I = James. I may be an amnesiac and may forget my name. (As I will argue in the next chapter, this should not be taken to rule out that identification of the subject is involved in self-ascription, but rather that it may be that self-identification is a type of expert skill that is extremely reliable.) In addition, one’s ability to engage in action also depends significantly on one’s I-thoughts. We do not say that one is able to self-ascribe one’s actions unless one is also able to affix to one’s actions, “I am doing such and such.” If we could not engage in self-ascription, then our bodily movements would be nothing to us (Kant (1781); McDowell (1996, p. 90)). They would seem to us to be alien—

someone else's actions, mental or bodily. In this sense, I-thoughts are pivotal for making one's mental happenings and bodily happenings into full-blooded actions.

I-thoughts also have import for affection or emotion. For example, if I think, "I just found out that I am a widower," then this has import for my emotional life in a way that "James just found out his wife died" does not. I-thoughts have an immediate import to one's emotional responses, in the sense that one (generally) has care and concern for objects of response that bears on one's existence. For example, if I think "I am being chased by a bear," then the emotional response of fear is prompted; however, that is not necessarily the case when I think "James is being chased by a bear." In addition, in order for emotions to have impact on my perception and action, I need to be able to think "I thoughts" about my emotions in such a way that it enables those emotions to be pervasive in my experience. For example, if I am to be said to perceive a prey animal and flee from a prey animal, then it is required that think "I am afraid of the prey animal."

So, what is the significance of this discussion about the relations between I-thoughts and perception, action and emotion? The key claim about self-ascription being made here is that self-ascription has an existential import. As Kant (1781, § B422-23) discussed in the First Critique, 'the I-think' is an existential concept; in Kant's terms it adds no further content to the categories, but instead plays a functional role in our mental economy. This will become central to the discussion below of Sellars' interpretation of the I-think as well as playing a role in my argument for a particular view of self-identification in Chapter 5. This is seldom discussed in the contemporary literature, and it is an important point to highlight. When we consider the ascription of experiential predicates in general, we need to account for a certain pervasive phenomenon, namely the

existential import of I-thoughts. We account for that existential import in terms of the functional role that such thoughts play in our mental economy. Now that we know more about what we mean by “I-thoughts,” we want to know how self-ascription based on I-thoughts is possible. In order to capture how self-ascription based on I-thoughts is possible, I want to begin by providing an analysis of its structure.

Before I do, I want to register a forewarning: I do not intend to discuss what a self is. If what I say below is to be related to the putative problem of self, then I am instead concerned with accounting for the structure “self-”, namely, what makes possible self-ascription of experiences. The structure “self-” should be thought to be a placeholder (as the I-think is for Kant) for *whatever is theoretically necessary to make sense of self-ascription*, since it is arguable that self-ascription requires the I-think. I will mark the placeholder structure with “I-think” to remain consistent with Kant’s usage. But, otherwise, ‘self’ is not one of my words and you will not find it in the dissertation. I am not concerned with the phenomenal self, the metaphysical self, the phenomenological self, or anything of the sort. In fact, my arguments here are perfectly consistent with the non-existence of *the self*.

I will now argue that there are a few desiderata that such a structure has to meet. (1) *The functional desideratum*: the structure has to account for the existential import that self-ascription has for experience in general (as outlined above). It is called the ‘functional desideratum,’ because the I-think enables one’s experiences to have a functional unity that would not be possible without the I-think. (2) *The compositionality desideratum*: the structure has to account for productivity of self-ascriptions. For example, an account of self-ascription should explain how if one is capable of thinking “I

am tired,” then one is also capable of thinking other thoughts “I am sleepy” and “I am hungry” and so on, for an idealized infinite number of predicates (productivity). The structure also has to account for the systematicity of self-ascriptions. For example, it must provide for an account of how those self-ascriptions form a system of connections between amongst themselves, for instance, if one is capable of thinking “I am tired” then one is capable of thinking “I am sleepy” and “Rip Van Winkle is sleepy” (systematicity). (Notice that the formulation does not presuppose that self-ascription meets the generality constraint, but rather that one’s account should be able to describe, explain and predict the canonical compositionality properties— productivity and systematicity— and as will be argued below the generality constraint best describes, explains and predicts this phenomenon.) (3) *The acquisition desideratum*: it has to be something that a creature can acquire in the normal course of development. *The acquisition desideratum* has already been discussed in the chapter on methodology, where I discussed Bermudez’s acquisition constraint. This desideratum will become especially important in the assessment of accounts of joint engagement in Chapter 6.

the subject-predicate structure of self-ascription

As I said above, I will argue that self-ascription has a subject-predicate structure. There are positions that agree with this assumption and positions that disagree with it. I consider positions roughly based on a helpful taxonomy given by Brie Gertler (2008) about self-knowledge, but it can equally be extended to self-ascription. I begin with a taxonomy so that it is clear that the first premise of the constructive argument does not depend upon a particular type of self-ascription, but equally well applies to most extant theories of self-ascription. Those that accept that self-ascription has a subject-predicate

structure are ‘Observation models’ which come in direct observation (Descartes, Chisholm (1977; 1981), Chalmers (1996: 221-2; 2002), Lehrer (2002)) and indirect observation (Locke (1689), Armstrong (1981), Lycan (1996)) varieties and ‘Rationality models,’ which come in constitutive rationality (Shoemaker (1994) and regulative rationality (Brandom (1994), Moran (2001)) varieties. I consider these views— not by illuminating all the details— but by considering what each account’s analysis of a self-ascription would be. After considering these views, I move on to the challenge provided by Humean Eliminativism and Neo-Expressivism.

The direct observation model suggests that when I think, “I am seeing a sheep,” I am observing an experience through introspection. This account argues that “I am seeing a sheep” actually has the content “I am directly observing myself as a subject seeing a sheep.” The direct observation model provides a mind’s eye account of self-ascription and suggests that self-ascription is a matter of our experiences being self-presenting through a type of direct mental perception. The subject of the self-ascription has her states presented to her in an immediate and direct way. Are direct observation theorists committed to the idea that self-ascription has a subject-predicate structure? On this model, the subject of self-ascription is “the I” that is introspecting a mental state. Given that the subject of the self-ascription is that to which the experiential predicate is presented and the presented object is a mental state, then the subject-predicate structure is accepted on this model. The subject of introspection is doing the observing and the predicate is what is being observed. Even if the subject perceives her experiences directly, the subject is nevertheless distinct from those experiences in the structure of the

observation. So, the direct observation theorist agrees that self-ascription has a subject-predicate structure.

Another observation model is the indirect observation model of self-ascription. Locke is perhaps the first to utilize an indirect observation model of self-consciousness: the “source of ideas, every Man has wholly in himself... And though it be not Sense, as having nothing to do with external Objects; yet it is very like it, and might properly enough be call'd internal Sense” (1689, § II.i.iv). For this account of self-ascription, “I am seeing a sheep” actually has the content “I am indirectly observing myself as a subject seeing a sheep.” This account of self-ascription argues that we have a special faculty of internal sense or inner sense. For this model, self-ascription is higher-order monitoring of one’s experiences. According to this notion of self-ascription, being aware of one’s experiences involves the positing of higher-order representations that indirectly capture the content of self-ascriptions. Are the indirect observation theorists committed to the idea that self-ascription has a subject-predicate structure? The model posits both a monitoring mechanism (even if the monitor is not presented) and the object of the monitoring (even if the monitored state or process is not presented). On the first order level is the experiential predicate itself, for instance “...seeing a sheep” and on the higher-order level “I am seeing a sheep.” To this extent, therefore, this view of self-ascription is committed to self-ascription having a subject-predicate structure, since within the self-ascription itself, there is a distinction between the subject and predicate.

The constitutive rationality model suggests that the capacity for self-ascription is constitutive of rationality, namely self-ascription plays a pivotal role in presenting an account of the constitutive ideals of rationality. In the constitutive variety, what makes

one a rational being is that being's capacity to self-ascribe one's experiences, most centrally one's actions. For this account of self-ascription, "I am hugging a sheep" actually has the content, "I have reason to think I am hugging a sheep." Being rational requires being capable of self-ascription and self-ascription is understood in terms of the capacity for the use of the first person singular pronoun in talk and thought about one's experiences. According to Shoemaker (1994), there is no observation necessary, whether direct or indirect, on the rationality model. Instead, self-ascription functions as a necessary accompaniment for the rational reporting of an experience. Are constitutive rationality theorists committed to the idea that self-ascription has a subject-predicate structure? In order to make sense of the agent herself offering reasons for her experiences, this would seem to be so. For instance, the agent needs to be able to distinguish between the bodily movements of hugging a sheep and the self-ascription of her agency of hugging a sheep. Thus, the content in question has to have a subject-predicate structure.

The regulative rationality model suggests that what makes one a rational being is that person's tendency to take up entitlements and commitments of endorsement for one's experiences. On this model, "I am seeing a sheep" actually has the content, "I am committed/entitled to think "I am seeing a sheep"." On this model, when one self-ascribes an experience, one takes up entitlements and commitments to the attitudes in the experience. For instance, when one says, "I am feeling sad," one avers the experience by endorsing the content of the experience. So, when one thinks "I am seeing a sheep," one is merely marking the content of the experience by saying that one is responsible for it. Is the regulative rationality model committed to the idea that self-ascription has a subject-

predicate structure? In order to make sense of the rational subject being capable of self-ascription, the regulative variety must be capable of determining that the subject of the self-ascription is something that is avowing, entitled or committed to the content being self-ascribed.

This taxonomy of views illustrates how each view is committed to self-ascription having a subject-predicate structure. I now must turn to those accounts that do not share the idea that self-ascription has a subject-predicate structure. The tendency to reject the idea that self-ascription has such a structure can be situated in intellectual history in Hume's claim (1978, § I.4.vi) that there is no subject of self-ascription. In order to answer Hume's eliminativism, I need to argue that the I-think has a subject predicate structure. But, Kant's argument for the synthetic unity of apperception has already done so, so I will contrast Hume's view with a version of Kant's argument in the B deduction. My reason for considering Hume's eliminativism is that if there is no subject of experience (even formally, because subjects are merely fictional bundles of perception), then self-ascription cannot have a subject-predicate structure. In the foregoing, I will provide an analysis of Humean Eliminativism, and respond to Humean Eliminativism via Kant's argument concerning the I-think.

When Hume reflected on self-ascription, he considered whether we are presented to ourselves. Hume's argument in the Treatise is that we do not have a legitimate idea of self, because we do not have an impression of a self. This is based on Hume's epistemological framework. Hume's epistemological framework for his skepticism about the idea of self derives from a legitimization procedure for ideas. Hume argues that ideas are either complex or simple. If an idea is complex then it must be reducible to simple

ideas. If an idea is simple, it must be derivable from a simple impression. If there is no corresponding impression for the simple idea, then the simple idea and any complex ideas that depend upon that idea are illegitimate. This enables Hume to eliminate those words, concepts or ideas that do not have adequate empirical criteria. In general, this legitimization procedure is very effective in keeping us from making claims using concepts that are not tied to empirical conditions. What about when it is turned on the idea of self?

Hume argues that the idea of self does not have a corresponding simple impression in the following famous passage:

“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. When my perceptions are remov’d for any time, as by sound sleep; so long am I insensible of *myself*, and may truly be said not to exist... I may venture to affirm of the rest of mankind, that they are nothing but a bundle of perceptions, which succeed each other with an inconceivable rapidity, and are in perpetual flux and movement” (252).

I take Hume’s argument to conclude that the feeling of connection between our perceptions— Hume actually means something rather general like ‘experiences’ or ‘experiential predicates’— leads one to imagine a fictional entity that we refer to with ‘I.’ The idea is that given that we only introspect experiences, the subject of self-ascription must be the product of those experiences. For Hume, self-ascription is based

upon a fictional description of a subject rather than a real subject itself. There is no corresponding object or referent for 'I'. The 'I' is a fictional entity that should be contrasted with the 'I' as an empirically real entity. Given Hume's lack of discovery of an impression, we have no proper epistemological basis for believing in the existence of the entity that may serve as the reference of "I." There are several contemporary views that take Hume as an ally in this idea. For instance, Daniel Dennett's (1991, p. 412) narrative notion of a self is a contemporary example. Also, Thomas Metzinger (2003; 2009) also presents a fictionalist account of the phenomenal self.

I agree with Hume and Humeans that there is no phenomenal self, as I have already said above, neither the dissertation nor the argument depends upon commitment to *the self*, whether phenomenal or noumenal. However, it is one thing to deny that we have experience of a phenomenal self, and another to deny that there is a subject/predicate structure of self-ascription of experiences. The point is that it does not follow from the successful rejection of the self as a legitimate metaphysical object that there there is an equally successful rejection of a structure of the I-think that determines self-ascription. And, I think Kant (and Sellars' (1970) interpretation of Kant's I-think) provides a way to make this distinction clear. Kant's implicit response to Hume is to argue for the need to appeal to a formal or structural posit called 'the I-think' in the transcendental deduction. Let me pause here to elucidate this argument.

The transcendental deduction begins with the idea that we possess self-conscious thought and experience, meaning our thought and experience is experience for individual subjects. As I discussed above, self-consciousness is a Moorean fact. According to Kant, self-conscious thought and experience implies objective thought and experience. This

suggests that the idea of a synthetic unity of apperception implies that there is a subject of experience that grasps a world (that is to take ‘experience’ as a success term). Objective experience is only possible assuming subsumption of objects under the categories.

Therefore, the categories are necessary for experience of objects. I will put aside the argument for the categories as such. That is not my focus here. (However, in a paper on Kant’s Schematism, I argue that there are important connections to be made between Kant’s schematism and Gareth Evans’ account of the object-dependence of thought.) Instead, what is implicit in this argument about the existential category of the I-think?

If we suppose that Kant is responding to Hume’s skepticism about the self, then how does the I-think respond to that skepticism? Kant distinguished between the empirical I-think, which might be the self as it is introspected in experience, and the transcendental I-think, which is the transcendental posit that helps to explain what makes subjective experience possible. What is the transcendental I-think? The transcendental I-think is the necessary accompaniment of judgments: “the I think must be able to accompany all my representations... [otherwise the representation] would not belong to the subject” (1781, §B132). The I-think is “a condition of the possibility of cognition of objects” (A354). It is the “he or it (the thing) which thinks... the transcendental subject of the thoughts = X” (1781, §A346/B404). It is important to notice that Kant does not provide a phenomenal description of the I-think, because the I-think is not something that has an empirical content. It is in this sense that a Kantian about the I-think as a transcendental posit is fully capable of agreeing with a Humean skeptic about the phenomenal self. The I-think is merely formal, so therefore it makes no commitments about the phenomenal self. However, that does not mean that the I-think is totally empty,

either. Strawson makes this point in The Bounds of Sense.

P. F. Strawson (1966, p. 102) makes a similar claim in the following about Kant's notion of self-ascription:

“...in Kant's terminology, that the possibility of ascribing experiences to a subject of experiences and hence the possibility of self-ascription of experiences requires that there be some “determinate intuition” corresponding to the concept of a subject of experiences; or, substituting the latter terminology for Kant's, we may say that this possibility requires that there be empirically applicable criteria of identity for subjects of experience. In actual practice this condition is satisfied by the fact that each of us is a corporeal object among corporeal objects, is indeed a man among men.”

Although philosophers and psychologists usually operate with the notion of a private, mental self that roughly parallels Hume's treatment of the self, the I-think is made possible by our being empirical, material, corporeal existence. As a matter of historical concern, I think this claim is something that Hume might have discovered. I want to pause to comment on Hume's legitimization procedure, as I discussed it above. Later in the same section of the Treatise, Hume considers by what relations our concept of self might be produced. As always, he turns to resemblance, causation and contiguity as the basic relations (Pitson (2002, p. 37–38)). However, he drops contiguity, saying it “has little or no influence in the present case” (260). According to Strawson, the notion of contiguity between oneself as an object and other objects, or between oneself as a body and other bodies, is basic to our understanding of our presentation of ourselves to

ourselves (as Cassam (1994, p. 141–42) has argued). What I am suggesting is that if Hume had considered contiguity, rather than rejected it, he might have discovered that the I-think is dependent upon impressions of the bodily self as a physical object among physical objects. (Cf. Quassim Cassam's (1994, p. 102)). This, of course requires much more argument than I can provide here, but Cassam's Self and World provides the first steps in the direction of bolstering this position, but without noticing this key mistake in Hume's account. I have not argued for this claim, but merely present it as a springboard for further interpretation.

The point of this contrast between Hume's skepticism about the phenomenal self and Kant's transcendental argument for the I-think is to suggest that while Hume is correct that there is no phenomenal self, the very possibility of experience requires the posit of an I-think. In addition, this I-think might be legitimated through one of Hume's avenues of legitimatization, namely contiguity. The claim is that the body/world structure is pivotal to self-ascription, for it provides the subject-predicate structure to self-ascription. One cannot ascribe predicates to oneself as a subject unless one as a prior condition can distinguish oneself as an embodied subject distinction from other material objects. It is in this sense that self-ascription must have a subject-predicate structure, because there is a material self-world distinction as its empirical basis. I will return below to the interpretation of Kant's I-think via Sellars when I discuss self-ascription in more detail. However, in this section, I have successfully argued that the Humean expressivist does not present a counterexample to the claim that self-ascription has a subject-predicate structure.

There is another view that might be taken to be a counterexample to the claim that self-ascriptions possess a subject-predicate structure. The next view I consider is Neo-Expressivism, for instance in Dorit Bar-On's account of self-ascription. I argue that Neo-Expressivism faces three objections, each of which are diagnosed as failures to notice the subject-predicate structure of self-ascription. One recent account of Neo-Expressivism is Bar-On's (2004) Speaking My Mind. I will not delve into the rich and interesting account presented in that book, but focus on the concern here about the form and structure of self-ascription. Bar-On considers avowals such as "I am seeing a sheep," "I am feeling hungry" (phenomenal), "I am thinking about my pickup" or "I hope the Yankees win" (intentional) as self-ascriptions. I wish to put aside Bar-On's account of the epistemology of self-ascriptions. My concern is strictly with what she takes to be the structure of self-ascriptions. "Semantically speaking, an avowal will typically identify an individual—the speaker—and ascribe to her a certain condition—being thirsty, feeling scared, hoping that...,etc" (2004, p. 8–9). For Bar-On's account, self-ascriptions have the same truth-value whether the self-ascription is thought via first person singular reference, say, "I am feeling scared about the approaching apocalypse" or via names like "James is feeling scared about the approaching apocalypse." It becomes clear that Bar-On's Neo-Expressivism is unique in considering self-ascriptions to be truth-evaluable.

One way to approach the details of Bar-On's Neo-Expressivism is through her analysis of Moore's Paradox. According to Bar-On, both "It is raining" and "I believe that it is raining" express our beliefs. However, the latter expresses the belief self-ascriptively. If one were to say, "It is not raining, and I believe that it is raining," then one would express a true proposition. But, the true proposition is just an extremely

unusual statement. As Brueckner (2005) has put it, “Though there is no contradiction between the two conjuncts, there is an expressive conflict, according to [Neo-Expressivism]. My utterance of the first conjunct normally expresses my belief that it is not raining, while my utterance of the second conjunct normally expresses my belief that it is raining. So [Neo-Expressivism] explains the oddity of the Moore-paradoxical assertion” (2005, p. 10).

So, when I think, “I am feeling hungry for tacos” I am expressing two things. The first is my thought that I am feeling hungry and the second is my mental state of feeling hungry. Bar-On calls this the Dual Expression Thesis (2004, p. 366). The Dual Expression thesis suggests that an avowal such as “I am feeling hungry for tacos” expresses both the predicate of the self-ascription and the activity of self-ascribing itself, however the activity of self-ascribing does not commit us, Bar-On thinks, to the subject-predicate analysis of self-ascription. Therefore, on the Neo-Expressivist view, self-ascription does not have a subject-predicate structure, because self-ascription does not involve the application of an experiential predicate to a subject, but instead a dual expression of the self-ascribed experience. I will now discuss three objections to this particular account of self-ascription. Two of my objections are based upon Brueckner’s (2005) sagacious critique of Bar-On.

The first objection is that the Dual Expression Thesis cannot be maintained for negative self-ascriptions, such as “I do not want guacamole.” Bar-on responds to this problem by suggesting that one has a mental state of wanting not to have guacamole. However, as Brueckner points out, this will not work for self-ascriptions such as “I do not believe that there are exactly 57 avocados in the bodega,” because one is not expressing

that the proposition that “there are exactly 57 avocados in the bodega” is false. This forces Bar-On to admit that there are self-reports of one’s experiences. To do so puts into question the Dual Expression Thesis, because it seems like Bar-On is forced to accept that self-ascription does have a subject-predicate structure.

The second objection is that the Dual Expression Thesis cannot capture two important phenomena of self-ascription. The first is that sometimes we do not express our self-ascriptions whether in action, non-linguistic or linguistic. For example, I may think, “I am hungry for tacos,” but not engage in any expressive action, non-linguistic or linguistic. As I discussed in the chapter on methodology, to ignore this goes against Bermúdez’s key insight about self-consciousness. However, my self-ascribing my being hungry for tacos has import for my perception, for instance, I am more likely to hear the taco bell on the corner. The contents of our self-ascriptions, however, still need to be understood as having some import for our experience in general, given the functional desideratum. The second is that the Dual Expression Thesis suggests that the activity of self-ascribing may be what constitutes the mental state in question. Since when we self-ascribe “I am seeing a sheep,” we are equally expressing both the self-ascription of the sheep and the seeing of the sheep, we are not in a position to make out a difference between expressing veridical self-ascription and expressing non-veridical self-ascription. In both of these cases, the diagnosis is that Bar-On has not considered self-ascription to have a subject-predicate structure.

The problem with the Neo-Expressivist’s account of self-ascription is that it does not except the expressing/reporting distinction. Let me pause to say a bit about that distinction. When one says, “I am hungry for ice cream” one is doing two things. One is

expressing one's self-ascription of the mental state of being hungry, on the one hand, and one is reporting upon the mental state of being hungry on the other (Sellars (1956, pp. 98–106)). For Sellars, experience contains claims that a subject may or may not endorse, and self-ascription plays the role of reporting upon the claim that one experiences such and such (I take up Sellars' Kantian account of self-ascription below). The endorsement of experiences is best understood in terms of the reports upon such experience, rather than expressions of such experiences. For example, a self-ascribed experience of being hungry, involves it seeming to one that one is hungry, rather than its merely being the case that there is a causal relationship between the mental state and the expression of it by the subject. The problem with Neo-Expressivism is that it fails to recognize this pivotal distinction between expressing and reporting. However, if Neo-Expressivists do so, they must concede the point that self-ascription has a subject-predicate structure, because reporting just is applying an experiential predicate to oneself as a subject.

In this section, I have argued that self-ascription must have a subject-predicate structure. But, so far I have not aligned my view with any particular account of self-ascription. However, I am not committed to doing so in order to make the constructive argument for *the Symmetry Thesis*. Instead, I am merely concerned with the structure and content of self-ascriptions. It should be clear, however, that the account of self-ascription I favor is inspired by Kant's view in the following quote: "All my representations in a given intuition must stand under the condition under which alone I can ascribe them to the identical self as my representations, and thus can grasp them together, as synthetically combined in an apperception, through the general expression I think" (1781, § B138). Self-ascription is achieved through a synthetic process, an activity of synthesizing one's

experiences as one's own.

2. Self-Ascription and Subject-Predicate Structure

I am concerned with a certain kind of self-consciousness, self-consciousness as the self-ascription of experiential predicates. According to what I have argued above, self-ascription of experiential predicates has a subject-predicate structure. For instance, "I am seeing a sheep" ascribes the experiential state of seeing a sheep to a subject. Self-ascription is the ability to ascribe experiences to oneself. The self-ascription of experiences involves thinking about one's experiences and attributing such experiences to a subject.

Let's look at an example to make this more clear and precise. Suppose that I am thinking about Meteora. When I think "I am thinking about Meteora," I am self-ascribing a content. When one is experiencing such and such, is one always self-ascribing one's experiences? Or is self-ascription a special type of mental process, such that one may be experiencing, but without self-ascribing one's experiences? If the former, then experiences are always self-ascribed. However, it should be clear that one of the major imports of philosophical psychology in the last 200 years is that we may entertain all sorts of states and processes without ever self-ascribing them. Perceptions, actions and emotions may be entertained in various ways without one having to think about them at all. So, what type of thinking is involved in self-ascription? To answer this question, we have to consider what ascription is in general.

Ascription is thinking about an object by applying a predicate to it. For instance, to ascribe the property of obesity to a child is to think "This child is obese." The structure of the ascription is a subject-predicate structure, in which "this child" is the

subject and "...is obese" is the predicate. The content of the ascription is the predicate of the ascription, for example, "being obese" plus whatever content is characterizing the subject, for example, "being a child." Ordinary predicate ascription such as this is not reflexive, for one does not need to think of the child as itself, but merely think of the child. However, self-ascription is reflexive, given that self-ascription is ascription of an experiential predicate to oneself as oneself. Again, it should be stressed that to say that self-ascription is 'reflexive' does not mean that it is 'reflective,' meaning that I need to occurrently reflect on my mental states or processes; nor does it mean that self-ascription is non-reflective, pace Gallagher (2005) and Zahavi (2007), because such a negative characterization is too broad to describe, explain or predict the phenomenon in question.

According to this account, self-consciousness is grounded in thought, because self-ascription involves the application of experiential predicates to oneself. One might object that this puts my project at odds with the phenomenological tradition according to which self-consciousness isn't dependent upon thought, but instead is dependent upon forms of perceptual awareness or action awareness. As I mentioned above, Gallagher (2005) and Zahavi (2007) both suggest, following in the footsteps of Husserl and Merleau-Ponty, that self-consciousness is grounded in a form of pre-reflective or non-reflective awareness. But, that my account of self-consciousness is based in thought and puts me at odds with the phenomenological tradition is an asset, not a liability, because I am not forced to subscribe to a special form of awareness to account for self-consciousness, which would leave self-consciousness mysterious. Although I do not address this in the dissertation, my model of self-consciousness aims to be reductive and naturalistic in the sense that it accounts for self-consciousness in terms of thoughts. And,

insofar as there is a perfectly naturalistic account of thought available, then self-consciousness does not require appeal to special forms of awareness. The way I approach thought is to consider ascription of predicates in general.

Ascription in general is the application of some or other predicates to a logical subject. In the case of thinking “This child is obese,” the logical subject is filled in by an obese child (assuming the ascription is true). How we come to recognize and acknowledge that there is such an obese child is a different question from its existence as a logical subject of ascription. The latter issue is a matter of the logical structure of ascription; whereas, the former issue is a matter of the content of the ascription. However, when we ask, what makes possible the ascription, then we have to be able to identify and re-identify the content of the ascription. For instance, focusing on the subject of the ascription, if one were to point to a child C_1 and say, “This child is obese” and then when asked, “Which child?” given that the classroom is filled with children. One might point to a different child C_2 (though that child may be qualitatively identical) and repeat “This one!” In this case, although the latter claim may be true (assume that all the children in the classroom are obese), the individual has failed to ascribe the property of obesity to the original child, because he has failed to possess the ability to identify and re-identify the original child in question.

This motivates a requirement of ascription in general. In order for an individual to be capable of ascribing some property or other to a subject, namely, a particular, that organism has to be able to identify and re-identify the particular in question (Strawson (1959: Chapter 2)). The measure of the success of ascription, in general, however, is that the ascription of the predicate to the logical subject tracks the particular. In my response

to Humean Eliminativism, I suggested that self-ascription has to depend upon some mode or way of determining the logical subject of self-ascription, what I will call ‘the sense of an I-thought.’

In order for this to be the case, then such a creature would have to be able to meet a condition that I call the ‘identification condition’: in order for a creature to self-ascribe, that creature needs to be capable of identifying and reidentifying the individual to which those predicates are ascribed. I should say here that this identification, while it may sometimes be conscious, is mostly a type of non-conscious identification. Here, I follow Kant in thinking that there are forms of non-conscious thought that provide a structure to experience. The identification condition presupposes a certain assumption about self-ascription, namely that the ability to self-ascribe experiences requires a certain degree of generality. But, this assumption has not been argued for. I pause here to reflect on what types of generality might be involved in self-ascription. It will become an issue whether the generality involved is the generality of the predicate or the generality of the subject.

Let’s return to an example, say “I am thinking about Athens.” The logical subject of the self-ascription is “the I,” the logical subject of the predication. The predicate of the self-ascription is “...is thinking about Athens.” There are two principles of self-ascription that will be taken up in my argument for *the Symmetry Thesis*: (SP) the subject principle (this will later be called ‘the Persons principle’, but for now SP is the more general contrast with PP) and (PP) the predicate principle. Let’s begin with the latter to articulate the kind of generality that might be involved in self-ascription. PP concentrates on the predicate of the self-ascription as central. The predicate principle of self-ascription suggests that the type of generality involved in the self-ascription is determined

solely by the content of the predicate, "...is thinking about Athens." In the case of "I am thinking about Athens" both thinking and Athens are the content, and whatever generality is required for ascribing such contents is by extension required in ascribing the predicate in self-ascription.

Let's consider the subject principle. SP concentrates on the subject of self-ascription as central. The subject principle suggests that the type of generality involved in the self-ascription requires determination not only of the content of the predicate, but also of the sense and reference of the subject of the self-ascription. In the literature on self-ascription and other-ascription, there is a tendency to focus only on the arguments for *the Symmetry Thesis* that rely on PP (I will discuss two historic accounts of the self-other relation in Chapter 6: the theory-theory and the simulation theory.) However, that is not the only means to the conclusion. While it is usually assumed that arguments for the *Symmetry Thesis* inevitably and irrevocably tied to the Wittgensteinian concerns about private language, the Strawsonian concerns about criteria of identification, or the Evansian concerns about verification conditions for senses, my unique argument for *the Symmetry Thesis* provides a different account. As I discussed in the introduction, focusing on the PP is not my strategy. However, I do need to approach these arguments and discuss what I take to be problematic about them, before I put them aside.

Let me pause briefly to consider the types of arguments for *the Symmetry Thesis* by focusing on the content of the predicates. There are arguments in the literature for *the Symmetry Thesis* that focus on particular predicates being symmetrical. To do so, I need to define 'symmetrical.' A predicate is symmetrical only if (1) the sense of the predicate is constant across first, second, and third personal uses; (2) one cannot apply the predicate

to oneself unless one grasps condition (1). There is a class of symmetrical predicates that could enable us to argue for *the Symmetry Thesis* based on this definition of symmetry. For instance, Dan Zahavi (2007: Chapter 6) discusses the problem of others minds and then highlights the continental philosophers approach via the robust concept of Empathy, or in German “*Fremderfahrung*.” What I want to highlight here is that the argument for *the Symmetry Thesis* based solely on this “empathic approach” only works for the self-ascription of certain experiential predicates and not others.

For instance, pride and shame are usually considered to be predicates that have a deep interpersonal content. For example, when I think “I am proud of my accomplishments” or (which is more likely) “I am ashamed of my failures,” then my self-ascription is constitutively linked to other subjects. Let’s call this the ‘interpersonal predicates thesis.’

The Interpersonal Predicates Thesis (IPT): there is a subset of experiential predicates that are interpersonal, and as such require a level of generality that involves their being used interpersonally.

The argument is that one cannot entertain interpersonal predicates unless one can ascribe those interpersonal predicates to others. I will not be arguing for IPT, but instead, I am merely highlighting IPT to discuss an argument for *the Symmetry Thesis*. These predicates are symmetrical predicates because they require generality across all uses: in order to ascribe pride or shame, one must distinguish between being proud (or ashamed) of oneself, on the one hand and being proud (or ashamed) of an other, on the other hand. The predicates of “pride” and “shame,” then, have a constant sense across those uses. In order to be said to grasp the sense of the interpersonal predicates, one must recognize this

fact. On the desert island, Solo (remember our banished and isolated child) would not be able to self-ascribe pride or shame, because without other individuals to be proud or ashamed of him (and he proud or ashamed of them), condition (1) could not obtain and neither would condition (2), because Solo would not be able to grasp condition (1).

All this might be true, but does it establish *the Symmetry Thesis*? It may be for these predicates, but if it does, it is relatively uninteresting, because the kind of self-ascription involved in pride and shame is already deeply interpersonal. Compare the claim, “the concept of marriage requires (at least) two individuals.” While this is no doubt true, it is an uninteresting fact that cuts no philosophic ice. Analogous with this, while it may be true that interpersonal predicates are taken up in our emotional lives are symmetrical in this way, this does not mean that all our experiential predicates are symmetrical predicates. Symmetry does not transfer from interpersonal predicates to all predicates, because other experiential predicates may be different. The tendency is to argue for a general claim about the intersubjective status of self-consciousness from the interpersonal nature of these predicates.

However, the argument for the general claim cannot go through. From the fact that interpersonal predicates are symmetrical, it does not follow that all experiential predicates are symmetrical. While it seems correct to say that the predicates of pride and shame are pivotal for our emotional lives, our interpersonal emotional lives cannot exhaust our subjective experience. For example, there are other experiential predicates—perceptions, actions, thoughts, beliefs desires— that are not *prima facie* symmetrical. I am not arguing that they are not symmetrical. Rather, I am merely suggesting that there

are significant differences between the class of interpersonal predicates and the non-interpersonal predicates.

Let's take an example of the latter. Merely by focusing on the predicate of seeing a sheep, one cannot extract from that predicate the fact that predicates of perception are symmetrical, because the content of seeing a sheep seems to be something that can be entertained without meeting the conditions on symmetry specified above. Merely by focusing on the predicate of agency, say, throwing a rock at a target, one cannot extract from that predicate the fact that predicates of agency are symmetrical. Merely focusing on certain predicates of emotions, for instance, feeling distraught, one cannot determine that that predicate is symmetrical. One may not have to grasp that the content of perception, action, and emotions are constant across all uses of the predicates involved in self-ascriptions of such experiences. The reason that thinkers have these intuitions about the contents of such experiences is that it is plausible to think that Solo in the desert island possesses perceptions, actions and emotions, so it seems equally plausible that he could think of himself as possessing such mental states. I am not arguing that this is the case, but just elucidating the important idea that the conclusion that can be reached from the interpersonal predicates thesis cannot be extended to the conclusion about other types of predicates.

The predicates of perception, action and emotion do not seem to be symmetrical predicates. In order to make that argument independent of the interpersonal content of the predicates, one would have to argue for what I will call (following Bermúdez (1998, p. 12) 'the constancy-of-sense thesis.' There are arguments that suggest that *the Symmetry Thesis* follows from views about the nature of predicates, in general, not just

predicates that imply interpersonal relations, such as the IPT suggested. I now consider that thesis, which I call ‘the constancy-of-sense thesis.’

The Constancy-of-Sense Thesis (CST): experiential predicates have a constant sense across first person, second person and third person uses, and as such, require a level of generality that involves reference to subjects other than oneself.

The constancy of sense thesis is that experiential predicates— no matter what the taxonomy of those predicates— have a constant sense. This can be captured in Strawson’s claim: “the dictionaries do not give two sets of meanings for every expression which describes a state of consciousness: a first person and a second-person meaning” (1959, p. 99). This is a droll way to express the idea that all experiential predicates have constancy of sense. The thesis is that in order for predicates to reach the needed level of generality, they have to be applicable in practice to arbitrary individuals. For instance, if the predicate “...is seeing a sheep” did not have a constant sense, then it could not be applied by a single individual at different times, nor could it be applied to different people at the same time, both of which are required by *the compositionality desideratum*. (This has been argued by Wittgenstein in Philosophical Investigations, but I will not pause to discuss Wittgenstein’s private language concerns.) Again, the compositionality desideratum suggests the structure of self-ascription has to account for productivity and systematicity of self-ascriptions. CST, however, faces a few objections and ultimately does not support *the Symmetry Thesis*. I will consider two objections to CST.

The first objection to CST is that different contents seem to have different degrees of constancy. For instance, "...is seeing a sheep" or "...is throwing a rock at a target" or "... is feeling distraught" seem rather constant across first person and second person uses. While the self-ascription "I am seeing a sheep" and the other-ascription "You are seeing a sheep" differ in mode of presentation of the experiential predicate, they do not differ in sense. I see the sheep from my point of view or perspective; you see it from yours. But, these differences in perspectives do not make a difference to the sense of the predicate, or so it might be argued. Regardless of who is seeing the sheep, the intentional content of the seeing is constant across both our self-ascriptions. Similarly, the self-ascription, "I am throwing a rock at a target" and the other-ascription "You are throwing a rock at a target" may differ in mode of presentation of the subject, but do not differ in sense. I throw the rock from my point of view; you throw it from yours. But, these differences in agency do not make a difference to the sense of the predicate. Similarly for the case of "I am feeling distraught," *mutatis mutandis*. I am not here arguing for CST, but instead merely describing some motivations to accept it. Roughly, this claim is that the predicates involved above all have the same truth-conditions across first-, second-, and third-personal uses, so they are symmetrical predicates.

However, even if one accepts this, it might be argued that there are a set of experiential predicates that are strictly first personal in our phenomenology, which I will call 'first person singular predicates,' relying on the notion of first person singular as something restricted to oneself. For instance, pains, itches and tickles seem to be such first person singular predicates, and *prima facie*, seem to lack the generality required by CST. The self-ascription, "I have a headache" and the other-ascription "You have a

headache” differ in mode of presentation of the subject, but, many have argued that they differ also in the sense, given that ‘headache’ as it is ascribed to me from my point of view always refers to my headache, and to no one else’s. This would, of course, contrast with Wittgenstein’s argument against a private language, which he describes as “The words of this language are to refer to what can be known only to the speaker; to his immediate, private, sensations. So another cannot understand the language” (1953, § 243).

Let me pause to consider that argument. The basic conclusion is that a language that is assumed to be in principle not understandable to the individual that invented its terms is impossible. The argument for this conclusion is that such a private language would also, in principle not be understandable to the individual. The putative private language user could never determine the senses of the predicates of that language. Can this argument be used to show that pains, itches and tickles have a constant sense across first-, second- and third-personal uses? I recognize that there are private language considerations that are meant to show that pains, itches and tickles do have a constant sense. I am more inclined towards this position, than towards the position that there are genuinely ‘first person singular predicates.’ However, I cannot resolve this dispute here, as that would take us too far afield. My hypothesis is that it may be that constancy is on a spectrum, where first person singular predicates such as pains, itches and tickles are *prima facie* less constant, whereas interpersonal predicates such as pride and shame are *prima facie* more constant, and therefore symmetrical. Regardless, the general CST is at least brought into question by this objection from first person singular predicates, and as such, we cannot infer from CST that *the Symmetry Thesis* is true.

The second objection to CST is that it depends on verificationist assumptions. To say that a predicate has a constant sense requires that it be applicable in practice to arbitrary individuals. For example, if I think “I am seeing a sheep,” then I would have to be able to apply the experiential predicate of seeing to an arbitrary individual. However, that would require that one be capable of verifying that the other was seeing a sheep. The constancy of sense thesis suggests that in order to apply a predicate to oneself, one must be able to carry out a method or procedure of verification that determines the object to which the predicate applies. However, this assumes a verificationism about senses that is implausible. I already accepted the idea that verificationism about meaning in general is difficult to defend. So, unless one can discharge this verificationist assumption about the senses of the predicates of self-ascription, the constancy of sense thesis cannot be maintained.

I have highlighted two theses, the interpersonal sense thesis and the constant sense thesis, both of which are assumed to be essential for arguing for *the Symmetry Thesis*. However, they are not able to succeed in arguing for that thesis for the reasons I discussed. The problem with these arguments is that both IPT and CST assume that *the Symmetry Thesis* can only be argued through a consideration of the content of the predicates. I have no doubt that given more work on these theses, an argument for *the Symmetry Thesis* might be defended. But, since there is such a difference across the continuum of experiential predicates from those that seem to have contents that are interpersonal and those that seem to have contents that are first personal, I will not go this route. The reason for this is that *the Symmetry Thesis* can be argued by focusing instead on the subject side of self-ascription.

My strategy, thus, involves sidestepping the debate about the predicate side of self-ascription and instead discusses the possibility that there might be a generality constraint upon the subject side of self-ascription. For instance, if one thinks “I have a headache” it is not the predicate we should be focusing on, but instead the subject of the self-ascription. It is puzzling why no one has taken up this strategy directly. So, the focus on the strategy is on the question of whether self-ascriptions like “I have a headache” involve ‘generalizing the subject.’ My claim is that whether or not an organism is able to entertain an experiential predicate does not tell us whether or not an organism has the ability to self-ascribe in general. It may be that there is a special ability to self-ascribe experiential predicates as opposed to other predicates, but without an argument for this special status, the ability in general needs to be accounted for.

Remember that any account of self-ascription has to account for how we are productive and systematic in our self-ascriptions, according to the compositionality desideratum. But, if we restrict our account to interpersonal predicates or to predicates that seem to have a constant sense, then we cannot meet the compositionality desideratum. It was *Strawson’s insight* about self-ascription that we have to account for how we self-ascribe a complete set of predicates to ourselves, not merely a subset. In Strawson’s case, the insight had to do with being equally concerned with the larger set of “person predicates” as we were with “mental predicates.” And, he thus fundamentally changed the conversation about whether mental predicates or bodily predicates were more central for a solution to the mind-body problem. Effectively, I am suggesting that we should bring about a change of subject in our account of self-ascription.

We should focus on the subject of self-ascription, rather than the predicates of self-ascription in order to argue for *the Symmetry Thesis*. As such, it will focus on the question of how to interpret the I-think that I discussed above. Unfortunately, Strawson does not offer an analysis of the subject side of self-ascription as effectively as Sellars did. For this reason, I turn now to a Sellarsian interpretation of the I-think as a basis for my account of the subject of self-ascription. Sellars both offers a way to make clear how the account of the I-think is central for discussing the subject-predicate structure of self-ascription and offers a way to articulate the I-think that supports the claim that it must meet the generality constraint.

the Sellarsian interpretation of the I-think

The account of self-ascription that I am offering here is a functional account, because it considers self-ascription to be a function of the understanding in general, one that can be theorized about independent of questions of content. Again, as I mentioned in the introduction, I will not enter the confused debate about whether self-ascription is conceptual or non-conceptual. This type of functionalism is inspired by Sellars's interpretation of Kant's I-think. I will now discuss Sellars' account of the I-think in order to make clearer the type of functional account I have in mind and to provide a further articulation of the basis for the claim that self-ascription must meet the generality constraint.

According to Sellars, Cartesian psychology wrongly asks the question, "what sort of being is an I?" (1970, 5§3). Transcendental psychology, however can be studied, provided we focus on necessary conditions— we should stick to a minimal characterization of *the I that thinks*. As I argued in Chapter 1, we are concerned with that

which makes possible self-ascription, rather than an inquiry into the nature of the self. Kant rightly pointed out that Cartesian psychology looks into the nature of the necessary condition for consciousness, and mistakenly infers about its properties. Hume makes a similar mistake above, inferring from the fact that self-ascription is the ascription of experiential predicates, so if there were a self it would be something experienced. Instead, transcendental psychology should inquire into the logic of “the I,” which is best understood by illuminating a function, which is the role of “the I” in our thought in general.

Sellars outlines the role of “the I” as an “*unrestricted* principle in the philosophy of mind” (7§7). He suggests that it unifies the manifold without implying that representations are had in or possessed by a subject. It is for this reason that Kant should not be read as holding an ownership or authorship, or source view of self-ascription, pace (Peacocke (in conversation)). Sellars has historically, been the key figure to argue that the I-think of Kant’s Paralogisms (1781, §§ A341-405/B399-432) is the “I” of apperception in Kant’s Deductions (especially §15-§27, B129-B169) are best understood in terms of a functional description of abilities to do such and such. The best evidence for this view is in CPR, A400. Once we understand that the world is collected under a number of forms, that the veridical (objective) representation of the world is made possible by the forms of sensibility (space and time) and understanding (the judgments and categories), then we can more easily reflect upon the significance of “the I” as the transcendental judgmental activity.

But, to understand that the I-think does not bring with it a need to analyze the content of “the I”, we should recognize that Kant points out, the “I-think” has to be

“assigned to” (A341/B399) the categories without altering them. There is no new content added. As will become significant for the discussion of the predicate principle in Chapter 5, Hurley (1998: 60–66) presents this point as an objection against views that assume a representational content of “the I” is required. In general, acts of thinking are best understood under the banner of judgment-activity, or “thought-activities,” which involve a singular and basic synthesis of all the forms of thought-activity under the categories and what Sellars has suggested is an über-category— the synthetic I-think. However, importantly for my purposes here, Sellars rejects the claim that is often made by Kant interpreters that Kant held an empty subject view of the ‘I.’ This view is usually motivated by appeal to the passage in the Paralogisms in which Kant says that the I-think is “an empty representation I, of which one cannot even say that it is a concept” (A345-6/B404).

Sellars’ reasons for rejecting this interpretation derive from the suggestion that Kant cannot be arguing that there is an empty “bare particular” ‘I’ which is determined by its ascriptions or by its attributive predications. This is Strawson’s view of Kant’s account of self-ascription in Bounds of Sense (1966, p. 94ff), and a view that contrasts with Strawson’s own account of self-ascription in Individuals (1959, p. 99). Kant could not hold the view that attributes or accidents inhere in a particular representation ‘I,’ since he warns against using the category of substance in this way in the First Analogy (A182f). Also, just because the I-think never receives a determinate content, that does not mean that the I-think is either illegitimate or empty. Rather, as I will discuss below, it is filled in by its empirical conditions.

Sellars suggests: “the conceptual burden of the “proposition ‘I-think’” is carried by the verb ‘to think’” (9§14). This generally aligns Sellars with the tradition of Kant interpreters (for instance, Cassirer (1927) and Heidegger (1997: 55)) that give pride of place to the logical function of judgments in Kant’s metaphysical deduction. Kant begins the deduction with the structure of judgments (A80/B106). According to Sellars, Kant saw the categories as “concepts of functional roles of mental activity” (11§20). The claim being made about the functional role of self-ascription is independent of a characterization of the content of self-ascription.

The ability for self-ascription is really two distinct functional roles. There are two distinct abilities to self-ascribe, what I will call ‘the subject ability’ and ‘the predicate ability.’ The subject ability involves the ability to attach the I-think to one’s representations. The predicate ability involves the ability to have the experience in question. The former ability introduces no new content to the content of the predicate, but instead is a placeholder for the process of determining that one’s experiences are one’s own. Being able to apply predicates to objects in general enables the predicate ability. The ability to apply experiential predicates is merely an extension of that ability. The subject ability is an ability to single out the subject of those experiential predicates. Above, I discussed how Kant did not specify what is required of this ability. However, that does not mean we cannot.

I will articulate the subject ability inspired by the work of Strawson and Evans. The possession of the subject ability depends upon being able to do a variety of tasks. When the subject is capable of attaching an I-think, then that subject has to be able to rely on the various empirical conditions for that process. I now spell out those conditions.

While I cannot argue for these conditions, however, there are figures that have made arguments for these conditions.

The first condition is that the creature have a sensibility (Strawson 1959).

The second condition is that the creature locate itself in space and time (Strawson 1959)

The third condition is that the creature be able to orient itself by tracing out a line in space and time (Strawson 1959).

The fourth condition is that it be capable of distinguishing between appearance and reality (Russell (1995, p. 134)).

The fifth condition is that it know-which embodied object it is (Cassam (1994, p. 108)).

I include these five conditions here only to illustrate a picture of the function of self-ascription. The above conditions are discussed in the literature on self-ascription, however, a sixth condition has not been the focus of much discussion: *in order to think I-thoughts, one must be able to recognize and acknowledge other individuals like oneself*. It is the sixth condition that is my concern in arguing for *the Symmetry Thesis*. Now that I have outlined the ways in which self-ascription requires the subject-predicate structure, I now turn to the argument for the claim that self-ascription must meet the Generality Constraint.

3. The Subject-Predicate Structure and the Generality Constraint

The argument for the claim that self-ascription must meet the generality constraint is found in the work of Strawson and Evans in a cocoon form. It is one of my goals to allow this butterfly to emerge from the cryssalis. The argument depends upon

recognizing a point that is made by P. F. Strawson in the essay “Persons.” Strawson the elder argues that self-ascription of ‘a is F’ entails ‘someone is F’; “someone” is an indefinite particular; indefinite particulars require a definite particular to secure ascription. Therefore, if ‘a is F’ is to be self-ascribed, then an empirical criterion for a definite particular is required.

Let’s consider an example can bring this down to earth. If one thinks “I am seeing a sheep” then this entails “someone is seeing a sheep.” The latter embeds an indefinite particular in the subject place. In order for self-ascription to succeed the “I” needs to be affixed to a definite particular. This argument is implicit in the following: “It is a necessary condition of one’s ascribing states of consciousness, experiences, to oneself, in the way that one does, that one should also ascribe them, or be prepared to ascribe them, to others who are not oneself. ...the ascribing phrases are used in just the same sense when the subject is another as when the subject is oneself.” (1959, p. 99). While Strawson’s discussion here is not presented as an argument, I believe the above exegesis captures the argument that Strawson had in mind. However, there are many things that remain unclear and imprecise in this quote. To begin, it remains unclear what ‘in the way that one does’ means. But, thankfully, Evans (1982, p. 259) addresses this concern.

Evans argues, along similar lines as I have argued above that there is a primitive mode of presentation or way of being presented to oneself as oneself through an empirical mode of presentation (1982, p. 259). As I suggested above, the way of being presented to oneself is through determining the experiential predicate of self-ascription. This determination is achieved through a variety of empirical criteria. When we are

talking about the way that one self-ascribes we are positing a sense that explains what makes possible self-ascription, and in doing so, we are required to appeal to other individuals. This idea is pivotal for the argument for *the Symmetry Thesis*. In a recent work also, Strawson argues “being equipped to identify some states of mind in others is a necessary condition of being able to ascribe any states of mind to ourselves” (1985, p. 22).

Bermúdez summarizes and criticizes these embryonic arguments by Strawson and Evans (1998, p. Chapter 9). However, I do not think that Bermudez’s paraphrase nor his critiques are any good. In order to defend the idea that self-ascription must meet the generality constraint, I will defend the argument of Strawson and Evans against Bermúdez’s objections. By addressing his objections below, I effectively defend the second premise of the constructive argument: self-ascription must meet the generality constraint. Bermúdez critiques Strawson’s and Evans’s claim that self-ascription must meet the generality constraint. He focuses on P. F. Strawson and Gareth Evans’s defense of the claim in the following argument:

- “1. Self-ascription of psychological states in subject-predicate form is the paradigm form of psychological self-awareness.
2. Self-ascription of psychological states requires conceptual mastery of both subject and predicate.
3. The Generality Constraint applied to (2) requires the ability to generalize both the subject and the predicate.
4. Generalizing the predicate involves the ability to apply that predicate to arbitrarily distinguishable individuals.

5. Since the predicate is a psychological predicate, these arbitrary distinguishable individuals will include other psychological subjects” (1998, p. 234).

In Bermúdez’s objections to this argument he rejects the connection between 2 and 3 and also separately rejects 4 and 5. However, as I discussed in the introduction, Bermúdez supposes that a Strawsonian account of concepts and concept mastery is tied to an account of language and language mastery. But, given that I have argued that the conceptual/non-conceptual issue is orthogonal to the debate about self-ascription, it still may be the case that the generality constraint follows, even once this assumption is dropped. Representations in general may require the generality constraint whether those representations are non-conceptual or conceptual, so it does not serve as an objection to the argument against self-ascription meeting the generality constraint to argue: if concepts are tied to language and self-ascription is not conceptual, that therefore self-ascription does not meet the generality constraint. Notice that this does not require that I argue against the thought-language principle—the only way to analyze the capacity to think a particular range of thoughts is by analyzing the capacity for the canonical linguistic expression of those thoughts—that Bermúdez presents, because it is not a matter of analysis. Instead, the claim is that self-ascription still may meet the generality constraint, regardless of your account of the relation between thought and language. In addition, neither Strawson, Evans nor I would accept premise 4 and 5. Recognizing that to construe the argument this way is a misunderstanding of the Strawsonian position. And, this is pivotal to the dissertation, and it will play an important role in what follows. I will pause briefly to expound on this misunderstanding, but it will take up a good portion of Chapter 5.

Also, I made a distinction between the subject principle and the predicate principle of self-ascription. In Bermúdez's (1998) account of the conditions of self-ascription, he makes an assumption that will be relevant to the argument later. Bermúdez reflects on what exactly we apply such experiential predicates to. In general, we have the assumption that certain individuals are thinking individuals and certain individuals are not. And, in the case of experiential predicates, the nature of the predicates *qua experiential* constrains the types of individuals to which the predicate applies. However, I do not think that Strawson or Evans are committed to following Bermúdez in accepting this assumption, nor do I think it is a fair assumption to make.

To repeat what was said above, the predicate principle is the assumption that the logical subject of the self-ascription is determined by the content of the predicates. There is reason to question this assumption, however. While I agree that "minimal requirement for something to be a psychological subject is surely that at least some psychological predicates are true of it" (Bermúdez 1998, p. 235), it does not follow that what makes a subject of a self-ascription must be the predicates that are ascribed. This would make it impossible to understand for instance, how a psychological subject could also have non-psychological predicates ascribed to it, say that a psychological subject has a broken arm, is late for the meeting, or has just been swimming, all of which (and many others) are necessary. To think otherwise is to deny Strawson's insight that is presented in the essay "Persons," namely that we ascribe both mental and bodily predicates to the same individual particular, namely a person, rather than to a bifurcated individual made up of a mental and bodily parts. The logical form of self-ascription does not suggest the latter analysis, but instead, suggests that the subject principle best captures self-ascription.

The subject principle analysis of Strawson's and Evans's arguments from self-ascription to the Generality Constraint, however, concentrates on the subject of self-ascription as central. The subject reading focuses on the conditions that constrain what counts as a subject of self-ascription. Bermúdez's Lockean argument for the predicate principle is: (1) the logical subject of self-ascription is determined by the predicates; (2) the predicates describe one logical subject, but no actual or particular subject; (3) therefore, there is no particular subject, but only that subject as taken up in the self-ascription of experiences. Neither Strawson nor Evans are committed to the predicate principle, nor to the Lockean commitment to introspective experience as the only means by which to account for self-ascription. However, the argument from the Generality Constraint need not focus on such predicates, but instead given that self-ascription has a subject-predicate structure, should focus on the generalizing of the subject, as Strawson's and Evans' arguments both do. My argument against the alternative Lockean account of the empirical basis for self-ascription comes in Chapter 5. For now, however, I discuss the motivation for the claim that self-ascription must meet the generality constraint. What is the Strawsonian argument that self-ascription must meet the generality constraint?

I argue that because of its structure, the ability to self-ascribe has to meet a constraint called the Generality Constraint. Gareth Evans (1982, p. 75) defines the generality constraint as: if S is to think "a is F" then S would need to meet the following conditions: (GC₁): S would need to be able to think "b is F,"; (GC₂): S would need to be able to think "a is G" (for any concept G that S possesses which is not F). For example, is one is to think "This sheep is grey," then one would need to be able to think "This

parakeet is grey” (or something similar for any grey particular) and S would need to be able to think “This sheep is stinky” (or something similar for any concept that S possesses other than ‘grey’. The basic argument that self-ascription must meet the generality constraint is that if a subject is to be able to self-ascribe the experience “I am F,” then she must have two distinct abilities: (GC₁) she must be able to ascribe experiential predicates, e.g., “b is F,” “c is F,” to arbitrarily distinguishable individuals; (GC₂) she must be able to ascribe “I am G,” “I am H,” for any experiential predicates she possesses independent of “I am F.” My task is to approach Bermúdez’s two objections to the generality constraint.

The first objection depends upon arguing that self-ascription does not require a level of generality that Strawson and Evans suppose. Bermúdez objects that the Generality Constraint can be met by application of experiential predicates to oneself over time, e.g., “I was sad yesterday” and “I am sad today,” or by the negative application to inanimate objects, e.g., “That rock is not sad.” So, according to Bermúdez the generality constraint does not require generalizing to arbitrarily distinguishable individuals, but only to oneself at other times, or to other non-psychological things at the same time. This may be rephrased as the following claims:

Bermúdez’s claim (1): diachronic generality is sufficiently arbitrary for generalizing of the subject.

Bermúdez claim (2): synchronic generality is sufficiently arbitrary for generalizing the predicate.

I have a two-fold reply to this objection. Bermúdez’s claim (1) is clearly circular. In order to ensure that the same subject is self-ascribing “I was sad yesterday” and “I am

sad today” one must presuppose subject-identity. It seems as if Bermúdez is attempting to explain the generality merely in terms of a mere repetition of self-ascriptions. But, as Butler (1736, p. 100) is correct to point out in his critique of Locke’s account of self-identity, if one has a strictly psychological criterion of subject-identity, then one’s account of self-identity is bound to be circular. Assuming subject-identity does not provide one with an account of subject identification. Bermúdez’s claim (2), however is not circular, but it commits a category mistake. It commits a category mistake because it supposes that psychological predicates are the types of predicates that can be applied in principle to inanimate objects. First, this leads to an implausible form of panpsychism for one, which is absurd. Second, in order to say that one meets the generality constraint, one cannot presuppose that experiential predicates potentially apply to any arbitrary object, because that would remove their informativeness. Third, experiential predicates are limited to the lives of animals. While it is an open question whether squirrels entertain experiences, it is not an open question whether rocks do.

The second objection that Bermúdez presents is that Strawson’s and Evans’s argument above cannot capture the difference between synchronic and diachronic identity. Bermúdez (1998, p. 236) argues that they merely focus on synchronic identity: “It is unfortunate, however, that both Evans and Strawson take a very synchronic view of the satisfaction conditions for their respective [generality] constraints. Had they taken a more diachronic perspective, they would have recognized that a subject himself can, over time, provide a range of different occasions on which a predicate may be significantly affirmed.” Bermúdez is arguing that one can meet the generality constraint for generalizing the predicate by “ascribing, or being prepared to ascribe, predicates to

oneself at different times” (1998, p. 236). But, this illustrates Bermúdez’s interpretative confusion between the constraints on self-ascription, on the one hand, and constraints on continuity of the subject, on the other hand. Bermúdez is again presupposing the latter in order to argue for the former, because he must assume that at a time, one already meets the constraints on the continuity of the subject, meaning that one is the same subject over time. One cannot presuppose that one has the ability to ascribe predicates to oneself at a time in order to argue that one has the ability to ascribe predicates to oneself at that same time. So, the second objection is equally circular.

Given that I have addressed Bermúdez’s objections to the generality constraint, I now move on to three other objections. The first objection I outlined briefly above. That objection suggests that my argument for the generality constraint on self-ascription assumes an implausible account of concepts. The second objection is that this account of self-ascription cannot allow for the pivotal distinction concerning self-consciousness between “I-as-subject” and “I-as-object.” The third objection is that the picture of self-ascription cannot account for the fact that self-ascription is immune to error through misidentification.

The first objection is that I have merely assumed that if self-ascription has a subject-predicate structure, then it must meet the generality constraint. I have provided no argument for this conditional, except by presupposing an implausible account of concepts. But, since there is no real consensus on what a concept is, it is illegitimate for me to assume a particular notion of concepts in order to argue that the generality constraint follows. However, I want to make clear that the argument that self-ascription requires meeting the generality constraint follows from a claim about the structure of

self-ascription. The claim that self-ascription must meet the generality constraint follows from the claim that the structure of self-ascription is a subject-predicate structure. Insofar as a self-ascription has the constituent structure “{I} {am Φ },” then the ability to self-ascribe must meet the generality constraint. Any account of self-ascription has to account for the fact that the ability to self-ascribe is systematic and productive— to meet the compositionality desideratum— and that is all that is assumed here. There is no general account of concepts assumed. Without an assumption of constituent structure, one could not meet the compositionality desideratum.

The second objection is that if self-consciousness is understood in terms of self-ascription, and must meet the generality constraint, then my account of self-consciousness cannot capture the pivotal distinction between awareness of oneself *qua* subject and awareness of oneself *qua* object (Shoemaker (1968)). The objection suggests that because the generality constraint requires a kind of universality that is indicative of “I” ranging over objects rather than ranging over subjects. However, self-consciousness as self-ascription is a kind of awareness of oneself as a subject, awareness of oneself as a subject is distinct from awareness of oneself as an object, therefore, the implication of the account provided here is implausible. I will not meet this objection in this chapter.

Instead, it will take up the bulk of Chapter 3. I will argue in that chapter that the distinction between I-as-subject and I-as-object amounts to a dogma and is unclear and imprecise, it cannot be used to object this account of self-ascription. But, for now, I will outline what the distinction is supposed to be in order to elucidate this objection.

The distinction between I-as-subject and I-as-object is supposed to go back at least to Kant (1781). However, one recent way of capturing the distinction can be found

in Kriegel (2008). According to the party line about the I-as-subject/I-as-object distinction, there is a difference between the following sentences:

(S1) “I am smelling a putrid stench in New York City”

(S2) “I am aware of myself as smelling a putrid stench in New York City”

In the case of S1, I am aware of an object, the putrid stench in NYC. In the case of S2, I am aware of myself as an object, because I am aware of the person smelling the putrid stench in NYC. But, also in the case of S2, I am aware of myself as the subject of that perception of a putrid stench in NYC. It is this latter way in the case of S2, that I am aware *qua* subject. A quick way to make this distinction is to say that the latter way in the case of S2 uses ‘I’ to refer to the subject of experience and the former way in the case of S1 uses ‘I’ to refer to the object of experience.

However, while Kriegel may be correct that we use ‘I,’ ‘me,’ ‘myself’ and related terms in these two distinct ways, this does not support the view that there are two distinct kinds of self-consciousness that are based upon two exclusive modes of presentation of the self. This is what I will argue in Chapter 3, when I critique this position about the distinction between I-as-subject and I-as-object. In the case of S2, it is equally plausible to suggest that the ‘I’ used in the latter way, or I-as-subject way is nevertheless dependent upon the former way, or I-as-object. However, there is little reason to follow Kriegel in making a distinction between awareness of oneself as a subject and awareness of oneself as an object in a way that Cassam suggests that they are two “mutually exclusive modes of awareness” (1994, p. 4–5). I will present an account of self-ascription by way of the I-think that is functionally dependent upon empirical criteria. The argument for this is that we cannot attribute mental states to a subject unless we can attribute mental states to

a creature in general, for example, to a person, or a human being, or a creature possessing some properties or other. The I-as-subject use does not float freely from these empirical conditions. In chapter 3, I will address this distinction between I-as-subject and I-as-object.

The third objection is that if self-consciousness is understood in terms of self-ascription, and self-ascription must meet the generality constraint, then this account of self-consciousness cannot capture the fact that self-ascriptions are immune to error through misidentification, because it implies that self-ascription is dependent upon self-identification. I have two replies to this objection, which I will take up in detail in Chapter 4. In the first reply, I argue via a conceptual argument and in the second reply I argue via an empirical argument. The conceptual argument focuses on Evans's account of self-ascription as IEM. I think that there is an interpretation of Evans's account of self-ascription that should not be committed to IEM. To Evans interpreters, this will of course seem tantamount to rejecting Evans entire account of self-ascription, but I do not think this is so. I will argue that the phenomenon (assuming there is such a phenomenon) of immunity to error through misidentification (IEM) only applies to self-reference, and that any account of self-ascription would do better to avoid the commitment that self-ascriptions are IEM. The empirical argument is that in the case of self-ascription, both for ascription of psychological and non-psychological states, there are cases in which reference to the subject is subject to error. I, will, take up the reply to the third objection in chapter 4.

I have argued for the claim that self-ascription has a subject-predicate structure. I have also discussed the picture of self-ascription that I think fits with what I have argued,

although I am not committed to arguing for this account in order to make the constructive argument for *the Symmetry Thesis*. In the last section, I argued that self-ascription needs to meet the generality constraint. For the argument for *the Symmetry Thesis* to go through, much more needs to be said. I have to argue why even if self-ascription must meet the generality constraint, why it is that in order for one to entertain the I-think, one must be able to apply (or at least be capable of applying) that structure to others (or at least possible others). In order for one to be able self-ascribe experiential predicates, one needs to be able to apply the I-think to oneself and to others. The next chapter will focus on the premise that suggests that in order to generalize the subject, one must be able to identify oneself. Since the idea that self-ascriptions are IEM is taken to be an orthodoxy, in the next two chapters, I provide an argument for the claim that the ability to generalize the subject of self-ascription requires identification of the subject of self-ascription.

Chapter 3

“Think of a picture of a landscape, an imaginary landscape with a house in it. – Someone asks ‘whose house is that?’ – The answer, by the way, might be ‘It belongs to the farmer who is sitting on the bench in front of it.’ But then he cannot for example enter his house”

— Ludwig Wittgenstein (1958, p. 398)

Shoegenstein on Self-Ascription, Immunity to Error and the I-as-Subject

Contemporary accounts of the self-ascription of experiences are wedded to two basic dogmas. The first dogma is that self-ascription (both mental and bodily) is immune to error through misidentification relative to the first person singular pronoun “I” when used in the I-as-subject sense (henceforth, called ‘IEM’) (Shoemaker 1968, 1994; Evans 1982; Peacocke 1983; Cassam 1997; Moran 2001; Bar-On 2004; O’Brien 2007). The second dogma is the belief that there is distinction between the I-as-subject— or, awareness of oneself *qua* subject— and the I-as-object— or awareness of oneself *qua* object (Shoemaker 1968; Rosenberg 1986; Cassam 1994, pp. 56–61). For instance, it is assumed that the sentences “I have a headache” and “I have a broken skull” employ two distinct uses of the first person singular pronoun ‘I,’ the former is absolutely immune to error, while the latter is not so immune. Given these two distinct uses of ‘I,’ it is also inferred that there are two distinct modes or ways of being aware of oneself. The distinction in this case is between consciousness of oneself *qua* subject and consciousness of oneself *qua* object (henceforth, I will call the former ‘SC_S’ and the latter ‘SC_O,’ and the distinction between them ‘the SC_S/SC_O distinction’). In this paper, I urge that these dogmas are groundless.

I provide some historical background for the philosophical claim that IEM is based on the SC_S/SC_O distinction. The philosophical claim to be assessed is: self-ascriptions, say, “I am seeing a sheep” are IEM, because such self-ascriptions involve awareness of oneself *qua* subject. This is a common claim made in discussions about self-ascription. Sometimes, the claim is made explicitly; sometimes, implicitly. More often than not, philosophers appeal to Wittgenstein’s discussion in the Blue and Brown books of I-as-subject uses and I-as-object uses as a basis for the claim. For instance, Sydney Shoemaker has claimed that Wittgenstein held the view that self-ascriptions are IEM and held the distinction between awareness of oneself *qua* subject and awareness of oneself *qua* object. However, as I will argue below, this interpretation of Wittgenstein is actually a crossbreed of the views of Shoemaker and Wittgenstein, which I will call ‘Shoegenstein.’ I offer a few reasons to think that Shoegenstein is not Wittgenstein.

Apart from these historical concerns, I argue that if IEM is based on the SC_S/SC_O distinction, and there is no non-circular account of that distinction, then IEM is not based on anything. If the SC_S/SC_O distinction is not based on a principled difference, then we should understand self-consciousness as awareness of a subject as an object (Tugendhat 1986, p. 135; Cassam 1997), which would mean that SC_S and SC_O are not exclusive. As with all revolutions, the dawning of the new guard has its consequences, all of which are idyllic, obviously. One consequence of disposing of these dogmas is to allow for a positive naturalistic account of self-ascription: rejecting both of these notions removes the air of mystery surrounding the notion of awareness of oneself *qua* subject. Another consequence is to present an approach to self-ascription that stresses what I will call (following Avramides (2001)) ‘the lived position of the subject,’ which I urge is friendly

to Wittgenstein's later account of the subject of self-ascription.

First, I begin with a preliminary clarification of self-ascription. Throughout the paper, I will discuss the self-ascription of experiences. An example of self-ascription is "I am seeing a sheep." When one thinks, "I am seeing a sheep," one is ascribing to oneself the visual experience of seeing a sheep. One is not *merely seeing* in this case. In addition, "... seeing a sheep" is a state or process that is *for* the subject of the experience. If one thinks, "I am seeing a sheep," then we say that the predicate is the content of the experience of seeing a sheep and we say that the subject is the individual that "I" denotes in thought. The predicate has the content of visually experiencing a sheep. The subject has the content of that to which the experience is being ascribed. The structure of self-ascription is a subject-predicate structure, because it would not count as a self-ascription unless there were a distinction between the ascriptive predicate, on the one hand, and the subject to which that ascription is ascribed, on the other. As I will discuss below, this analysis of self-ascription neither requires that self-ascriptions are IEM nor that there is an exclusive distinction between SC_S or SC_O.

By way of a disclaimer, I should say that I will not be providing conceptual reasons and empirical evidence against IEM directly. I have said that I am concerned in this paper with the philosophical claim that the distinction between those self-ascriptions that are immune to error through misidentification and those that are not depends upon a distinction between awareness of oneself *qua* subject and awareness of oneself *qua* object.

Other than disposing of the above dogmas, a goal of the chapter is defending the 3rd premise of the constructive argument for *the Symmetry Thesis*. Defending the 3rd

premise of the argument for *the Symmetry Thesis*— the generalizing of the subject requires identification of subject of experience— requires that I make room for the idea that self-ascription depends upon a type of self-identification. In this chapter, the first step to defending this premise is to defend the idea that self-ascription is ascription of experiential predicates to an object. The claim is that identification (and re-identification) of a subject only really makes sense if that subject is also considered a material particular, namely an object. For example, if I think “I am seeing a sheep,” then if the I-thought involves self-identification, then the sense of the I-thought would need to have a material particular as its object. In order to argue that self-ascription is not IEM, I first need to argue that IEM is based upon or grounded on the subject/object distinction. In the last section, I conclude by considering some implications of my skepticism about the subject/object distinction for an account of subjectivity.

1. Wittgenstein on the Possibility of Error and I-as-Subject Uses

Who is responsible for the claim that self-ascriptions employing the subject uses of “I” are immune to error through misidentification relative to the first person singular pronoun? Most discussions of IEM begin with a reference to Wittgenstein’s discussion of I-as-subject uses and I-as-object uses in second part of the Blue Book (pp. 48–74). A few long passages from the Blue and Brown Books bear quoting in full:

“Now the idea that the real I lives in my body is connected with the peculiar grammar of the word “I,” and the misunderstandings this grammar is liable to give rise to. There are two different cases in the use of the word “I” (or “my”) which I might call “the use as object” and “the use

as subject". Examples of the first kind of use 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 the second kind are: "*I* see so-and-so", "*I* hear so-and-so", "*I* try to lift my arm", "*I* think it will rain", "*I* have a 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 error, or as I should rather put it: The possibility of error has been provided for..." (Wittgenstein 1958, pp. 66–67)

"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 neighbors. . . 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 is you who have pains?' would be nonsensical" (Wittgenstein 1958, p. 67).

It is plain that in neither of these quotes does Wittgenstein say that either self-ascriptions are absolutely immune to error through misidentification relative to the first person singular pronoun or that awareness of oneself *qua* subject constitutes a special mode of

self-awareness that is distinct from awareness of oneself *qua* object. What Wittgenstein does do is highlight two features of the grammar of “I”-use, a point that I think Wittgenstein was correct to underline.

Also, it should be noticed that the distinction is between two uses of the first person singular pronoun “I” or “my” and not between the first person singular pronoun, on the one hand, and other terms or classes of terms, say “he” or “she” on the other. This is a point seldom realized. Talk about “I” when discussing its essential indexical nature is different from talk about ‘I’-as-subject. One may reject talk of the latter, while still accepting talk of the former. What is the difference between I-as-subject and I-as-object? What does Wittgenstein want us to notice about the different implications of the I-as-subject uses and I-as-object uses?

It seems that when we use “I-as-object,” we use it in a way that speaks about the human body and its physical characteristics. When we use “I-as-subject,” we use it in a way that speaks about mental states and processes, sensations, thoughts, etc. In the I-as-object case, an object, namely a person is recognized and something is predicated of the person (notice, however, that Wittgenstein never says ‘identified’). In the I-as-subject case, there is “no question of recognizing a person.” But, as we will discuss below, it does not follow from the idea that there is “no question of recognizing a person” that no recognition of a person is involved. However, philosophers have not restricted themselves to what Wittgenstein says, but instead, have taken what he says as an inspiration. It is Sydney Shoemaker’s “Self-Reference and Self-Awareness” that is the first to give a clear and precise account of IEM.

2. Shoemaker on IEM and Self-Presentation “From the Inside”

Shoemaker (1968, p. 81) discusses self-ascriptions such as “I am feeling a pain in my neck.” Shoemaker does not think that self-ascriptions are totally immune to error in every respect. One might be in error about the object of the content of the predicate. For instance, it may be that I misidentify the body part, misidentifying my shoulder as my neck. One might be in error about the state that enters into the content of the predicate. For instance, it may be that I misidentify the experience that I am feeling, misidentifying an itch as a pain. One might be in error about the way or mode of entertaining the content of the experience. For instance, one might be imagining a pain in one’s neck and confuse this with feeling a pain in one’s neck. But, Shoemaker argues, even if all of these types of misidentifications are possible, since there is no identification in the case of the “I,” there is no possibility of misidentification of the “I.” There is a logical and absolute immunity to error through misidentification in the case of subject uses of “I.”

Shoemaker’s analysis of IEM is the following: “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” (Shoemaker 1968, p. 82). Self-ascriptions like “I have a broken neck” are subject to error, according to Shoemaker, because that statement could be analyzed by the following inference:

- (1) a is Φ ; “Someone has a broken neck”
- (2) I am a; “I am that someone”
- (3) I am Φ ; “I have a broken neck”

The claim is that in cases in which self-ascriptions are IEM, this analysis is not proper,

because the second premise is not necessary. Instead, in cases of IEM, no inference is necessary because one seems to have non-mediated and non-observational access to oneself and one's experiences. For example, if I think, "I have a pain in my neck," then there is no mediating or observation-based identification that determines that that experience is my own experience. According to proponents of IEM, it does not make sense to ask "Someone is Φ , but is it me?" in the case of self-ascriptions that are based on awareness of oneself *qua* subject. However, in the case of self-ascriptions based on awareness of oneself *qua* object, this might be something one could ask (although whether one actually does is another matter). For example, when I say, "I have a broken neck" it might be that I base or ground that claim on the perception of my twin's crooked neck reflected in a mirror. In such cases, self-ascriptions are subject to error, because there is an identification component involved in the subject expression.

Gareth Evans takes up this intuitive notion of IEM, following up on Shoemaker's analysis. Evans defines IEM with a slightly different emphasis than Shoemaker: "a judgement is identification-free if it is based upon a way of knowing about objects such that it does not make sense for the subject to utter 'Something is F, but is it a that is F?', when the first component expresses knowledge which the subject does not think he has, or may have, gained in any other way." (Evans 1982, pp. 189–190). For Evans, if judgements are identification-free, then they are IEM. But, several questions arise about this account. How do we determine what it makes sense to ask? About this question, no one has given an answer. Evans might respond that it does not make sense to ask the question because it is plainly nonsensical to ask it. However, if what I argue below is correct, then what is supposed to be nonsensical only *seems* so because we are gripped by

a theory of self-ascription. Insofar as there is a debate between proponents and skeptics about IEM, then this theory of self-ascription cannot be presupposed. Is there any argument or evidence that there actually is a special way of knowing about ourselves? Shoemaker provides an answer to this question.

According to Shoemaker, when self-ascriptions are IEM, they are based on a *special* form of introspective self-acquaintance. According to Shoemaker, my self is “accessible to me (itself) in a way in which it is not accessible to others, so that in knowing that what is presented to me is presented in this special way — from the inside, as it were— I would know that it can be nothing other than myself” (Shoemaker 1968, p. 87). Shoemaker makes clear (1968; 1994) that he does not consider introspection to be perceptual or observational in any sense, and my critique of his position does not assume that his account of introspection is perceptual or observational. What is important to recognize for the purpose of this paper is that Shoemaker identifies self-ascriptions that are IEM with self-ascriptions that are based or grounded in awareness of oneself *qua* subject: “...self-awareness of the sort we are concerned with, does not involve being presented to oneself as an object” (Shoemaker 1968, p. 88). So, according to Shoemaker, self-ascriptions that are IEM are dependent upon consciousness of oneself *qua* subject.

(There is a threat of this debate turning into a war about shifting the burden of proof. The proponents of IEM argue that because it seems like there is no identification involved in awareness of oneself *qua* subject, then there is no identification involved. The skeptics about IEM argue that there could be an identification involved, and in fact, it is merely an appearance that self-ascriptions are identification-free. The skeptic about IEM, however, is not committed to arguing that there are identifications involved in self-

ascription, but instead merely that there could be.)

At this point, it will be helpful to characterize the position that Shoemaker takes from Wittgenstein with a different name. I will baptize the family of philosophers that are committed to the philosophical claim that *self-ascriptions are IEM when based or grounded on awareness of oneself qua subject* as ‘Shoegenstein.’ What is Shoegenstein committed to with respect to self-ascriptions? Shoegenstein is committed to the following conditional: self-ascriptions of the form “I am Φ ” is IEM only if “I” is based on or grounded in consciousness of oneself *qua* subject. If one thinks, “I see a sheep” then, according to Shoegenstein, such a self-ascription is IEM only if ‘I’ is based on or grounded in consciousness of oneself *qua* subject. Why might there be this conditional relation between IEM and SC_S?

Consider the denial of SC_S, namely the denial of the idea that “I see a sheep” is based or grounded in consciousness of oneself *qua* subject. In this case, it would be based in the consciousness of oneself as an object. But, if this were the case, then such a consciousness of an object would have to involve recognition and identification. Insofar as that self-ascription involved recognition and identification, then it would involve the inference:

- (1) a is Φ ; “Someone sees a sheep”
- (2) I am a; “I am that someone”
- (3) I am Φ ; “I see a sheep.”

However, if this were the case, then such a self-ascription would be subject to error through misidentification, because (2) could be false.

One way to make this clearer is to consider a distinction that Shoemaker employs

between absolute immunity and circumstantial immunity. In cases of self-ascriptions in which there is no identification of the “I”, for example, “I see a sheep” there is no possibility of misidentification. There is an absolute immunity to error through misidentification in the case of subject uses. However, the object cases like “I am facing a sheep” have merely circumstantial immunity. There may be identification failures, for instance, in cases in which I look in a mirror to see myself facing a sheep.

Many philosophers inspired by Shoemaker have supported a regress argument concerning IEM. If all self-ascriptions were merely circumstantially immune to error, and every self-ascription is based or grounded on some other self-ascription, then we would quickly find ourselves falling into an infinite regress. Shoemaker accounts for this regress with the following description:

“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-based 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 (Shoemaker 1996, p. 211).

According to Shoemaker, then, at least some self-ascriptions must be absolutely immune. So, if we deny that a self-ascription is IEM, then we have to deny that that self-ascription is based or grounded in recognition or identification of a person, otherwise we run into this dreaded infinite regress.

According to Shoegenstein, therefore, IEM and SC_S are two sides of the same coin; if we are faced with one side, then the other side is always present. It is safe to say that philosophers of self-consciousness have taken this dubious coin to be basic to our understanding of self-consciousness, even going so far as to say that a distinctive *mark* or *feature* of self-consciousness is IEM and SC_S , as Uriah Kriegel (2008) has argued. These two ideas are assumed to be mutually reinforcing. In section 4, I will question whether the distinction between SC_S and SC_O can be legitimated. For now, I turn to the claim that Shoegenstein is not Wittgenstein.

3. Shoegenstein is not Wittgenstein

It has become commonplace in the study of Wittgenstein's work to remember to stress that Wittgensteinian philosophy is a type of philosophical therapy. Wittgenstein did not present theories or accounts, but instead assembled reminders to avoid philosophical conundrums. According to this anticonstructive method, the goal of philosophy is to deconstruct our assumptions that lead us to think theories or accounts are necessary. As he discusses in the *Tractatus*, "Language disguises thought. So much so, that from the outward form of the clothing it is impossible to infer the form of the thought beneath it" (1961, §4.002). In the case of self-ascriptions, Wittgenstein should be taken to hold the position that the superficial logical form of self-ascriptions gives rise to philosophical questions that the real logical form would not. "Most of the propositions and questions of philosophers arises from our failure to understand the logic of our language" (1961, §4.003). Many philosophers of self-consciousness that deserve to be called 'Shoegenstein,' however, have ignored Wittgenstein's anti-constructive leanings, and instead adopted his claims in the Blue Books as means for theorizing about self-consciousness. I do not wish to push this objection, however, since (1) it is controversial whether Wittgenstein's own method was fully anti-constructive and (2) whether taking inspiration from elements of Wittgensteinian philosophy requires one to commit to an anti-constructive method. I will, however, return to this point in Chapter 7 when I discuss the dissolution of the conceptual problem of other minds.

The first point in favor of the claim that Shoegenstein is not Wittgenstein is that is unclear whether Wittgenstein wanted his talk of the possibility of error and the distinction between the subject use of "I" and the object use of "I" to cut any philosophic ice. Shoegenstein has employed the distinction as a gouger to make points about self-

consciousness in the philosophy of mind. Thus, we have to ask about Wittgenstein, “What is the import of the distinction between I-as-subject uses and I-as-object uses?” Wittgenstein does say that we are *highlighting different functions of language serving two different human needs*. Thus, we might say that the I-as-subject uses are our means of expressing our experiences: “to say, ‘I have pain’ is no more a statement *about* a particular person than moaning is” (Wittgenstein 1958, p. 67) and that the I-as-object uses are our means of reporting upon ourselves as having or possessing states and processes. In the subject case, no object is referred to. In the object case, an object—a person—is recognized and identified and something is predicated of the person. But, Shoegenstein wants to do more than highlight functions of our language, namely two distinct uses of “I,” and the different pragmatic assumptions we make between expressions and reports. Shoegenstein wants those features to do heavy lifting in the philosophy of mind. In particular, he wants to be able to infer that there are two distinct forms of self-awareness—*qua* subject and *qua* object. It is unlikely that Wittgenstein would have joined him in this.

The second point is that Wittgenstein talks about error in the case of subject uses of “I” as “not being provided for” and Shoegenstein discusses error as being absolutely impossible. These two are *prima facie* different. To say that *the possibility of error has not been provided for* in the case of I-as-subject reference does not entail that I-as-subject reference is *immune to error*. Consider the following analogy. Take an example of a train running down a track. There are all sorts of aspects and features that make it the case that *the possibility of error has not been provided for*. The context of the rails, the turns being of such and such arc, the speed limit, the weight of the train, etc. It would be

correct to say that *the possibility of error has not been provided for*, assuming that context is one that is error removing. However, it does not follow that the train is *immune* to running off the track, because there are always influences outside of the context that could lead to error. When things are operating normally, within the context, errors do not arise. However, when things are not operating normally, the train is subject to running off the track. Maybe Shoegenstein would argue that self-ascriptions that are IEM relative to awareness of oneself *qua* subject are simply different in kind than things like trains in which the possibility of error has not been provided for because of the error-eliminating context. But, another feature of Wittgenstein's work on intentionality in general, is that in order for items of talk and thought to count as being intentional, then there have to be correctness and incorrectness conditions involved. However, if self-ascriptions are absolutely immune to error through misidentification, then the idea of correctness and incorrectness evaporates. So, it does not seem that Wittgenstein would have followed Shoegenstein down this track.

The third point in favor of the idea that Shoegenstein is not Wittgenstein is that Shoegenstein uses the distinction between I-as-subject uses and I-as-object uses to motivate a distinction between two ways or modes of appearing to oneself, one of which is subjective and the other objective. For instance, "I see a sheep" is based or grounded on the subjective way and "I am facing a sheep" is based or grounded on the objective way. However, Wittgenstein might have held the view that uses of terms and concepts enables us to reject what we might call "objective referentialism" about our talk and thought: "when we perceive that a substantive is not used as what in general we should call the name of an object, and when therefore we can't help saying to ourselves that it is

the name of an ethereal object” (Wittgenstein 1958, p. 47). But, instead of assimilating this reminder, Shoegenstein postulates *special ways or modes* of being aware of ourselves in the I-as-subject cases. However, if Shoegenstein wants to determine a way or mode in terms of awareness of oneself *qua* subject and awareness of oneself *as* object, then he would have to assume objective referentialism with respect to the latter. But, since Wittgenstein rejects that view *tout court*, then the contrast does not get off the ground in the case of uses. Thus, even if we allow the weighty interference from pragmatic features of language use to facts about mental phenomena, the contrast does not get off the ground in the case of modes of awareness either.

The fourth point is that Wittgenstein himself rejected the distinction between I-as-subject and I-as-object as being *even* conceptually distinct in his mature philosophy. The Wittgenstein of the *Philosophical Investigations* does not maintain that statements such as “My arm is broken” and “I am feeling arm pain” are disconnected in the way that Shoegenstein suggests. This is the gravest mistake. Two uses cannot be made to straddle a conceptual divide that cannot be unraveled; Shoegenstein depicts I-as-subject uses and I-as-object uses as two games in tandem, never overlapping, always unconnected. It is in this respect that Shoegenstein is stuck in the Blue Book characterization of the distinction between awareness of oneself *qua* subject and awareness of oneself *qua* object.

However, as Hans Sluga has pointed out, the formula for the resolution of this false bifurcation is that “an ‘inner process’ stands in need of an outward criteria” (Wittgenstein 2001, p. 580) (Sluga 1996, p. 337). According to Wittgenstein, “only of a human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; hears; is deaf; is conscious or unconscious” (Wittgenstein 2001, p.

281). These outward criteria are not something that are understood in terms of two exclusive modes of access to oneself, but instead of a single mode of access to the form of life that we engage in as living human beings. It might be objected that the private language considerations only apply to the predicate terms of the self-ascription. Namely, if one thinks “I have a pain,” then the private language considerations are concerned with the predicate ‘pain’ rather than the function of self-reference. However, in the quote from the *Philosophical Investigations* that begins this chapter, Wittgenstein does address self-reference directly, pointing out the difficulty in making sense of the function of self-reference. While he rejects a descriptivist interpretation of self-reference in that passage, this should not be taken to imply that Wittgenstein thinks that there is no reference of “I” (*pace* Anscombe (1994)). (For an interpretation of Wittgenstein’s account that diverges from Anscombe’s no reference view and presents an account of self-reference consistent with the account presented here, read McDowell’s (1998b) paper “Referring to Oneself”.) As I will discuss at the end of section 6, this supports the interpretation of Wittgenstein’s notion of self-ascription as embedded in the lived position of the subject.

4. No Non-Circular Account of the SC_S/SC_O Distinction

Upon what basis does Shoegenstein rest the distinction between those self-ascriptions that are IEM and those that are not IEM? The basis is rather thin; and certainly not as thick as to merit that self-ascriptions be assumed *de facto* IEM. David Rosenthal argues that the distinction is ultimately circular. Rosenthal discusses Shoemaker’s account of those self-ascriptions that are IEM as dependent upon I-thoughts that arise “from the special way we have access to our own conscious states” (Rosenthal 2005, p. 354). If I think “I am in pain,” then it is tempting to think that I cannot be wrong

because my consciousness of that pain via a subjective way guarantees that I am the thinker of the I-thought. Shoegenstein stipulates that access via a subjective way involves the guarantee that I am the thinker of the I-thought. But, as Rosenthal is correct to point out, “it is a substantive question whether the thoughts under consideration are immune to error through misidentification, and it begs that question to assume that all access that seems subjective is actually subjective in this stipulated sense” (Rosenthal 2005, p. 355). Apart from the historical concerns about whether Shoegenstein is Wittgenstein, I will take inspiration from Rosenthal and turn to the issue of whether Shoegenstein’s account of IEM is based on a distinction which itself does not have a basis or ground. I now turn to the claim that there is no non-circular account of the distinction between awareness of oneself *qua* subject and awareness of oneself *qua* object. There will be many philosophers in this section that come and go, but each deserves to be labeled with the moniker ‘Shoegenstein.’

4.1 “From the Inside” and “From the Outside”

I begin, however, with Shoemaker himself. As I said above, Shoemaker makes the distinction between SC_S and SC_O by distinguishing between awareness *from the inside* and awareness *from the outside*. It remains unclear what “from the inside” and “from the outside” is supposed to mean. When it comes to this type of distinction, no one should be satisfied with an elucidation of “from the inside” with “you know it when you feel it.” Because if I don’t feel what “from the inside” feels like, then I don’t know what self-conscious “from the inside” is. However, I am self-conscious notwithstanding my not knowing what self-conscious “from the inside” is.

The claim under consideration is that consciousness of oneself *qua* subject is best

understood in terms of awareness of oneself from the inside. Shoemaker does not expound in particular upon what this is supposed to mean. However, he does suggest (Shoemaker 1968, p. 84n4) that discussions of reflexive and non-descriptive self-reference can help us to understand what “from the inside” is supposed to mean. He does refer to Casteneda’s (1999, pp. 35–60, pp. 89–95) sagacious treatment of the irreducibility of first person singular reference. From this, we can add to what we said above that “from the inside” should be understood in terms of reflexive and non-descriptive self-acquaintance. First we consider the claim that SC_S should be understood in terms of reflexivity. One helpful discussion of Casteneda-style reflexivity is found in Robert Nozick’s chapter “Reflexivity.”

What is the real difference between those referents that are ‘reflexive’ and those that are ‘non-reflexive’? This might be made clearer by considering Nozick’s account of self-consciousness, since it is in this context that we get an explicit treatment of reflexivity. Nozick urges that “the I” involves reference to a self that is a reflexive relation (Nozick 1981, p. 110). About the reflexive self, Nozick argues that the self has a “distinctive reflexiveness,” because each of us has a care and concern for ourselves as ourselves. Nozick accounts for this distinctive reflexivity of “the I” in terms of “being reflexively self-referring” (Nozick 1981, p. 112). But, while it might be made transparent what non-reflexive reference is, since it is the garden-variety kind of reference, is it obvious what “being reflexively self-referring” involves? I cannot help but think that calling anything “reflexive” will invite a lack of clarity rather than make the phenomenon more transparent. I offer two objections to this type of account in terms of reflexivity.

The first objection is that in order to make sense of referring, whether it is merely self-referring— for instance, “James sees a sheep”— or reflexively self-referring— for instance, “I see a sheep”— we need to appeal to some object. An act of reference without a referent is not an act of reference at all. An act of reference, for example in the statement, “The Nothing noths” is not an act of reference to the Nothing. So, “being reflexively self-referring” must involve some type of objective reference on the side of the referent. The second objection is that a similar argument can be made about the agent of the self-referring. An act of reference without an agent of reference is not an act of reference at all. For example, an act of reference involved in “I am seeing a sheep” requires the subject of the act of reference to be an agent of the reference independent of the act. So, “being reflexively self-referring” must involve some type of objective reference on the side of the agent of the reference.

However, Nozick would argue that being reflexively self-referring is made possible by the fact that reflexively self-referring is referring *as reflexively self-referring*. But, how this can be anything other than running in place is lost on me. It is the phrase that enters into the *as reflexively self-referring* that we are hoping to understand. Reiterating the notion simply labels the problem, rather than giving us the answer we are longing for.

Consider the other claim that “from the inside” is to be understood as a type of non-descriptive self-awareness. So, when I think “I see a sheep,” I think about myself from the inside in a way that no description could adequately represent. In the famous example by John Perry (1979), an absented-minded professor is pushing a shopping cart around a supermarket. He notices that there is a thin trail of sugar on the floor, and says,

“That idiot is making a mess.” Suddenly, the professor realizes that it is *he himself* that is spilling the sugar, and says, “I am making a mess.” In the first statement, Perry is in fact referring of himself. However, simply being aware of oneself via a description is never sufficient. In the latter statement, Perry is referring to himself *as himself*. It is this non-descriptive relation, being aware of oneself *as such* or *as oneself*, that is necessary for self-ascription, according to Perry. However, this example makes two assumptions that are implausible.

The first assumption is that there are never any descriptions that *could* enable the professor to realize it was he himself that was making the mess. But, if the important phenomenon to capture is the fact that the professor removes the sugar bag from his cart, informs the supermarket employee of his messiness, or some other indication of his commitment and responsibility for his own actions, then it seems that there are descriptions that he could think that *could* serve this purpose. However, now we are foisted into an unending dialectic with Perry— a descriptivist proposes a description that would suffice and Perry constructs a “Perry case” that attempts to show that one could self-ascribe, even though they did not know that description. This is where the fly enters the fly bottle.

I offer two options for resolution of this dialectic. The first remedy is conceptual: imagine a logically perfect language in which every object has its own descriptive name. In such a language, a descriptive name for oneself would suffice to fix reference to oneself, since every object— including oneself— would have a descriptive name (Rosenthal (in conversation)). The second remedy is empirical. Consider the following question: “What is it reasonable to say we might not know or believe about ourselves

and still be able to act in the world?” It seems if we begin with a mass of descriptions— “...likes football,” “...bales hay in August,” “...is five foot seven,” “...is named ‘a’”, “...is here and now writing,” “...uses these eyeballs to see”— the closer we get to the empirical conditions upon I-thoughts, the more our intuitions will go against Perry’s and say that certain empirical descriptions are necessary to the operation of I-thoughts.

I have put into question the assumption that “from the inside” is non-descriptive. But how could we understand the distinction between non-descriptive and descriptive thought about ourselves, unless we have presupposed (wrongly) to have understood the distinction between SC_S and SC_O ? Inevitably, ‘non-descriptive’ is taken to mean ‘irremediably subjective,’ so we are caught in a circle. It is now clear that we cannot make sense of this cognitively significant difference between “from the inside” and “from the outside” unless we already grasp the difference between SC_S and SC_O . However, if this is so, then the locutions “from the inside” and “from the outside” cannot be the basis for that distinction. If this is the case, then the “from the inside”/“from the outside” distinction has to be put aside as a plausible account of the distinction between awareness of oneself qua subject and awareness of oneself qua object.

4.2 Intransitive Self-Consciousness and Transitive Self-Consciousness

Another way to make a distinction between consciousness of oneself *qua* subject and consciousness of oneself *qua* object has been via a distinction between intransitive awareness of oneself and transitive awareness of oneself. For instance, Uriah Kriegel (2003) makes a distinction between SC_S and SC_O by distinguishing between intransitive and transitive states. He contrasts two reports: “(a) x is self-conscious of her thought that p. (b) x is self-consciously thinking that p.” (Kriegel 2003, p. 103). The contrast

between these two is supposed to mark a difference between SC_S (which has the structure of (b)) and SC_O (which has the structure of (a)). Consider the distinction between the following thoughts: “I am self-ascriptively seeing a sheep” and “I am self-ascribing my seeing a sheep.” In the former case, it might be suggested that self-consciousness is intransitive, because it does not require the existence of an object. It is “peculiar” in the way that Wittgenstein originally suggested (Wittgenstein 1958, p. 158f: §15), under which no other content could take its place. The claim that is implicit in Wittgenstein’s notion of peculiar is that there is a *principled peculiarity* involved. According to this peculiarity construal of self-consciousness, the subject place of self-ascription is a place only for subjects, namely it cannot be held by objects; therefore, the *mark* of awareness of oneself *qua* subject is intransitive consciousness of oneself. The common ways to characterize intransitive consciousness of oneself is (1) through an adverbial treatment, which suggests that intransitive self-consciousness is “peculiar”— its principled peculiarity being inferred from the difficulty of turning modifications of verbs into nouns— and (2) through the claim that intransitive self-consciousness is awareness through a *non-relational way*.

Let’s consider an example to discuss the peculiarity construal in more detail. If I think “I am seeing a sheep,” in order to decide whether the better analysis is (a) x is self-conscious of his perception that p or (b) x is self-consciously perceiving that p, I would have to have some intuition or other about the latter, namely the intransitive interpretation. But, why should we admit that the adverbial treatment of the phenomenon is best? It can only be because we are supposing that *self-consciously seeing* captures something that we actually want to call, when we are being honest with ourselves,

subjectively seeing. How else can we communicate the difference between transitive and intransitive, except by talking about being aware of objects in the former case and not being aware of objects in the latter case? That is just what the distinction means normally. Therefore, the notion that the peculiarity construal can advance our understanding of the subject/object distinction is mired in swamp of circularity.

What about the idea that we could understand intransitive self-consciousness in terms of a non-relational way of being aware of ourselves? The notion of a non-relational way is that for the self-ascription “I see a sheep,” the kind of self-consciousness that is involved here is non-relational. But, if we suppose that “I see a sheep” is understood in terms of I-thoughts, and then that such I-thoughts are non-relational, then it just seems unclear what ‘non-relational’ is supposed to mean. One reason is that I-thoughts are not posited merely to account for our presumed direct and immediate awareness of ourselves, *pace* Shoegenstein’s best efforts. Our intuitions culled from introspection about how our experience of ourselves when we think “I see a sheep” might *seem* or *appear* non-relational, non-observational, not mediated. But, that *apparent* non-relationality cannot be used to infer that such I-thoughts *are* non-relational. And, apart from this skepticism about introspection, the positing of I-thoughts is relevant precisely because of the relational role and character of such thoughts. For instance, we need to describe, explain and predict why if I entertain the self-ascriptions “I see a bear” and “I am afraid of a bear,” then I also run from the bear. We want to understand how I-thoughts enable relations between our experiences such that they have import for action. I-thoughts are at least relational in this sense.

However, it is likely that Shoegenstein would respond that I am confused, because I do not understand what it means to say that I-thoughts are non-relational: “When we say that I-thoughts are non-relational, we mean that such thoughts do not involve relations to objects.” But, this just makes the circularity transparent. So, the intuitions that are marshaled to account for the difference between the transitive and the intransitive types are just intuitions about the distinction between awareness of oneself *qua* subject and awareness of oneself *qua* object. So, the distinction between the transitive and intransitive depends upon a prior understanding of the SC_S/SC_O distinction, and so cannot serve as a basis for the distinction.

4.3 I-as-Agent and I-as-Patient

Another means of capturing the distinction between awareness of oneself *qua* subject and awareness of oneself *qua* object is via a distinction between awareness of oneself as an active being and awareness of oneself as a passive being. For instance, when I asked my Grandfather at Thanksgiving to distinguish between the self-conscious subjects and the non-conscious objects in the room, he said that the people, the dog and maybe the fly on the wall would be the subjects and the dead turkey, the fireplace bellows, and the gravy boat would be the objects. When prompted for reasons, he “waxed Aristotelian” and suggested that he made this distinction quite naturally based on the assumption that the former are capable of being agents (actually, he slurred, “moving”) and the latter are not (“not moving”). This can seem like a rather natural way to distinguish between subjects and objects in general. It seems that subjects of self-consciousness are active, moving, and might even be entitled/committed to what they do. On the other hand, it seems that objects of self-consciousness— suppose for the sake of

discussion, we consider one's brain or body— are without the requisite dash of subjectivity, because they are passive, immobile, and not entitled/committed to what merely happens in them. In this sense, it seems that SC_S is agentic; SC_O, passive. There are several figures that have stressed the importance of agency for self-awareness. (For good examples, see Evans 1982, p. 207; Campbell 1994 and O'Brien 2007, Chapter 5).

One especially good one, however is McDowell's (1998b) paper, which takes up Evans' unfinished project and attempts to spell out how agency is pivotal to subjectivity. McDowell characterizes the kind of agency that is involved in self-consciousness in the following: "a potential for reflectiveness belongs to thought as such; "I" encapsulates that potential into a way individual things can be presented in the contents of thinkings" (McDowell 1998b, p. 141). For McDowell, there is a "specific agency exercised in [self-referring]" (McDowell 1998b, p. 142). There is a potential for active reflection that is agentic and is responsible for creating awareness of oneself *qua* subject. However, the patient side— one's thinking being governed mutually by the world and others in the world— is responsible for creating awareness of oneself *qua* object. McDowell's view is then similar to Nozick's Fichtean view discussed above, in which there is an agentic self-referring that is determinative of the subject. Insofar as it is the similar, the same objections apply to this view. However, apart from McDowell's Fichtean view of self-consciousness, if we focus on the agentic nature of self-consciousness, two objections to the agency view arise.

The first objection is that we often self-ascribe thoughts that do not have a sense of agency or action involved in them at all. There is nothing particularly agentic about our thoughts in general or our I-thoughts in particular. Often, thoughts like "The sheeps

will escape if the gate is open,” “Shut the gate,” or most importantly for our purposes “I left the gate open!” just strike us in a way that does not seem particularly agentic. If we take SC_S to be based on the agentic nature of self-consciousness, but, there are forms of self-consciousness that do not involve agency, then we have arrived at a dead end.

The second objection is arrived at by asking how we determine that a self-ascription is agentic versus non-agentic. For example, if one thinks “I am seeing a sheep” and one’s awareness of oneself as an agent is supposed to be the basis or ground upon which one says that one is aware of oneself *qua* subject, then we had better know what awareness of oneself as an agent involves. But, how do we determine that any event or occurrence happens with agency as opposed to without agency? And, in particular, how do we can distinguish between agentic I-thoughts and passive I-thoughts? One canonical way to distinguish between agent and patient is that bodily events or occurrences, for instance, unbuttoning one’s shirt are agentic *only if* one is the source of the action with a distinctive first-person sense of control (Marcel 2003). For example, if I am petting a sheep and I also have a sense that I am the source of petting a sheep, then it counts as an action. For instance, movements of an anarchic hand are not actions in that sense, because they lack this first-person sense of control.

If this is the case— namely, bodily movements count as action only if there is a first-person sense of agency in the action— then in order to make the distinction between agentic I-thoughts and passive I-thoughts, we need to make a distinction between those I-thoughts that are experienced as one’s own and those that are “nothing to one.” In the former case, these I-thoughts would be agentic. In the latter case, these I-thoughts would be merely happenings or non-actions. However, being able to make out this

distinction simply presupposes the distinction between the subject's own actions, on the one hand, and the happenings of or to an object, on the other. If this is so, then it seems clear that we cannot make out the SC_S/SC_O distinction based upon the agent/patient distinction. The latter seems to depend upon a prior understanding of the SC_S/SC_O distinction, and so cannot serve as a basis for the distinction.

4.4 *IEM and Not IEM*

Another way to make out the distinction between awareness of oneself *qua* subject and awareness of oneself *qua* object is to ground the distinction in the difference between those I-thoughts that are IEM and those that are not. Quassim Cassam supports this conditional. However, this is puzzling, because Cassam hopes to defend the view that self-consciousness is awareness of oneself as a material object (Cassam 1997). However, throughout, because of the pernicious influence of Shoegenstein, he considers self-consciousness as awareness of oneself *qua* subject as a material object. What does Cassam provide as a basis for the SC_S/SC_O distinction? Notwithstanding the fact that the distinction is pivotal to his project, the only account of it is by way of the distinction between IEM and \sim IEM. So, according to Cassam, if one's awareness of oneself is SC_S , then that awareness of oneself is IEM:

“Identification “necessarily goes together with the possibility of misidentification” (Shoemaker (1994: 87)), so awareness of something as an object cannot be the basis of statements that are immune to error through misidentification. This explains why there is a conflict between the idea that one is aware of the bodily self *qua*

subject and the idea that one is aware of it as an object among others in the world. If one is aware of it as an object among others in the world, then it must be possible for one to misidentify it, but if this is possible, then, *by definition*, one's awareness of it is not awareness of it *qua* subject" (Cassam 1997, p. 61; my emphasis).

According to this view, if I think, "I am seeing a sheep," then what makes that self-ascription involve awareness of oneself *qua* subject is that the subject reference with "I" is immune to error, and vice versa. If I think, "I am walking my sheep," then what makes awareness of oneself *qua* object is that "I" is not so immune. One might think that what makes awareness of oneself *qua* subject peculiar is this type of IEM. Other than Cassam's usage, this seems to be a common practice in the philosophy of self-consciousness when Shoegenstein holds sway.

However, it is immediately clear that the IEM/~IEM distinction cannot serve to make out the SC_S/SC_O distinction, since we are trying to make sense of the IEM as dependent upon the SC_S/SC_O distinction. We began with the philosophical claim that self-ascriptions are IEM, because they involve consciousness of oneself *qua* subject. Again, if we reject the latter notion, then IEM has no real import for self-consciousness. But, to base the SC_S/SC_O distinction on the IEM/~IEM distinction would land us with a vicious circularity. If it turns out that IEM/~IEM stands or falls with the SC_S/SC_O distinction, and the SC_S/SC_O distinction cannot be made clear and precise without IEM/~IEM, then we are forced to admit that these notions can provide us nothing but a ride on the philosophical sit and spin.

Notice that I have not addressed the issue of whether self-ascriptions are IEM can be made sense of independently of the SC_S/SC_O distinction. That is not my concern here. I have merely been concerned with whether self-ascriptions that are IEM can be distinguished from self-ascriptions that are not IEM by way of the SC_S/SC_O distinction. I now turn to some objections to my argument.

5. Objections

The first objection I will consider is that it is not clear why the claim that self-ascriptions are IEM or the SC_S/SC_O distinction are “dogmas” rather than just undeveloped claims or unclear phenomenology. For some Shoegensteinians, both the SC_S/SC_O distinction and the IEM/~IEM distinction are taken as hidden assumptions that are used to determine what is and what is not self-consciousness. Two cases spring to mind. The first is Uriah Kriegel (2008), who writes as if there is no way to capture the nature of self-consciousness unless we can capture the epistemic and semantic peculiarities involved in SC_S and IEM. The second is Joel Smith (2006), who argues that there are two forms of IEM and one captures what is peculiar about self-consciousness, so that form of IEM is correct. But, that has to assume that self-ascription is IEM. In these two cases, both IEM and the SC_S/SC_O distinction are wielded as means to rule in or rule out phenomena as self-consciousness or confirm/disconfirm theories of self-consciousness. In that sense, they are taken as dogmas. I have argued that either those distinctions are based in notions that are unclear or those distinctions presuppose a circle of notions that while they are mutually reinforcing, have little going for them “from the outside.” My goal has been to show in a piecemeal way that the SC_S/SC_O distinction needs to be based in more clear notions and needs to break out of the circle of notions.

The second objection is that the charge of circularity falls flat because there has to be some subject/object distinction assumed somewhere. On the one hand, Shoegensteins will argue, SC_S/SC_O distinction *does presuppose* the distinction between subjects and objects, because it distinguishes two ways of being self-aware in precisely those terms. On the other hand, SC_S/SC_O distinction is *not supposed to prove or vindicate* the subject-object distinction, because that is taken for granted. Shoegensteins say, “We take the subject-object distinction *for granted*, and so do not intend discussions of either IEM and \sim IEM, or SC_S/SC_O distinction to *vindicate* that distinction.” But, my response is that there is no independent subject-object distinction to be taken for granted. The family of notions comprises a global holism that embraces all of the kinds of distinctions that I discussed above (and probably others), and there is little hope of breaking out to discover a foundation merely by reiterating that that foundation is *taken for granted*. It is in the nature of foundations to be taken for granted, but if my arguments are correct above, then the foundations cannot be taken at face value.

The third objection is that there is no difference between self-ascriptions being absolutely immune to error through misidentification and maximally reliable at hitting their targets. The way I read the distinction that Shoegensteinians make between self-ascriptions being absolutely immune and merely circumstantially immune is that this is a distinction that is a matter of kind. The only way that Shoemaker’s regress argument that I elucidated above could go through is if this were so. However, the distinction between self-ascriptions being maximally reliable at hitting their targets and reliable is a matter of degree. As far as the naturalistic perspective supports, I think we should say that we actually do not know what is maximally reliable at this point. It is my claim that

accounting for the contingent reliability of self-ascriptions is the work of a theory of I-thoughts, and if I am correct, then the theory of I-thoughts will be an empirical theory. However, Shoegensteinians by comparison are wedded to saying that I-thoughts are absolutely IEM, which ultimately rules out presenting an empirical theory, because IEM is taken as an *a priori* datum, rather than a phenomenon that needs to be described, explained and predicted. As I will discuss below I would like to leave us open to a contingent reliability account that enables an avenue for an empirical account of I-thoughts.

6. Self-Ascription without the Dogmas

My skepticism about the SC_S/SC_O distinction should be transparent by now. However, often skeptics are good at arguing that their opponents are asking for loans based on faulty credit, when the skeptics themselves are secretly filing for bankruptcy. So a meta-skeptic might ask, “What is the cash value of rejecting the idea that self-ascriptions are IEM and dependent upon awareness of oneself *qua* subject?” If we reject the idea that self-ascriptions are IEM, then we are not making the claim that I-thoughts are absolutely immune to error. I suggest that it is simply a more naturalistic attitude to hold that self-ascriptions are maximally reliable relative to the first person. In this case, therefore, self-ascriptions of perceptions like “I see a sheep” might involve a maximally reliable transition from the perception of seeing a sheep to the self-ascription of that experience. According to this kind of account, self-ascriptions are dependent upon a form of self-identification, but that self-identification often piggybacks upon an artifact of linguistic grammar, the fact that “I” always refers to the speaker or thinker. Within many, even a heavily weighted most, the possibility of error would not be provided for.

However, that does not mean that such self-ascriptions are *absolutely immune to error*.

I want to pause to elucidate a similar account of self-ascription in the work of Elizabeth Fricker. Fricker discusses what she calls an Impure Artifact of Grammar theory of self-ascription. On her view, self-ascriptions are not understood as determined by “the concept-fixing conventions of ‘use’ alone” (Fricker 1998, p. 200). Nor is her account solely based upon the contingent reliability of self-ascriptions. She explains: “On the Impure AOG [Artifact of Grammar] theory, concept-fixing linguistic convention and empirical contingency play an intertwined and *inextricable* joint role in accounting for the reliability of Self-Ascriptions. Grammar contributes to the explanation of reliability, because the fact that a self-ascriptive utterance has a default status as correct is... a ‘bedrock’ fact of grammar; it is part of what fixes our concept of the mental state in question. But,... for any multi-criterial mental state concept, its viability depends on the contingency that the criteria hang together” (Fricker 1998, p. 200–201).

I am not concerned with Fricker’s account of self-ascriptions in general. I want to highlight an aspect of it that I take to be compatible with my skeptical account advanced above: self-ascriptions involve self-identifications. My extension of her comments to my account is the following: the artifact of grammar that the first person singular pronoun “I” is IEM provides a default status assumption that our self-ascriptions—whether they are linguistic or non-linguistic—do not depend upon identification. However, it does not follow from this artifact of grammar that we should infer that the actual empirical employment of I-thoughts are *absolutely immune to error*. After all, from a naturalistic perspective, insofar as there is a transition from an experience, on the one hand, to the self-ascription of an experience, on the other hand, then things could go

wrong. To think otherwise is to have a perversely optimistic conception of the human mind and brain, a place where Murphy's Law occasionally has application. The point is that there may be empirical criteria that underpin the thinking of I-thoughts and insofar as I-thoughts depend upon these empirical criteria, then such thoughts can be severed from the criteria.

You might ask as well what can be gained from rejecting the idea that self-consciousness is fundamentally awareness of oneself *qua* subject (other than obviating the circularity problems that I illustrated above). One reason is that maintaining such a view ultimately leads to an implausible thesis called 'the exclusion thesis,' namely that the self is not an object among other objects in the world. This cannot be argued in detail, although it is discussed by Cassam (Cassam 1997, pp. 9–21). Insofar as the self that is involved in self-ascription is understood in terms of that which one is aware of when one is aware of oneself *qua* subject, and the distinction between subject and object is taken to be exclusive, then the self is not an object among other objects in the world. There are a number of quotes of Wittgenstein's that seem to support this idea: "The I is not an object" (Notebooks, 80); "If I wrote a book called *the World as I found it*, it should have to include a report on my body, and should have to say which parts were subordinate to my will, and which were not, etc., this being a method of isolating the subject, or rather of showing that in an important sense there is no subject; for it alone could not be mentioned in that book" (Wittgenstein 1961, §5.631); "The subject does not belong to the world: rather, it is a limit of the world" (Wittgenstein 1961, §5.632).

However, just like Wittgenstein modified the bifurcation between subject uses and object uses in light of the private language considerations, Wittgenstein's considered

philosophy did not leave the Tractarian solipsism in place. Apart from this historical concern, I do not think that Shoegenstein should leave it in place either. For instance, consider the following quote from Shoemaker: “there is a perfectly good sense in which my self is accessible to me in a way in which it is not accessible to others” (Shoemaker 1968, p. 87). One of the benefits of rejecting the idea that consciousness of ourselves *qua* subject involves thought about ourselves via an irreducible subjective way, is that we can make sense of how we can also think about others. It is basic to our everyday experience that we accessible to each other in direct and immediate ways.

For instance, in a discussion of ascription of experiences, Sir Peter Strawson argues that the predicates that enter into self-ascriptions and other-ascriptions do not “have two kinds of meaning. Rather, it is essential to the single kind of meaning that they have, that both ways of ascribing them should be perfectly in order... the predicate gets its structure from the whole structure of the game” (Strawson 1959, p. 107). If I am correct that an account of self-ascription need not be wedded to the Shoegensteinian idea that self-ascriptions are IEM or that I-thoughts depend upon awareness of oneself *qua* subject, then we can recover from the idea that self-ascription and other-ascription are two distinct activities, mediated by inferences, imaginative projections, and theories about theories. Instead, we can make clear how self-ascription and other-ascription depend upon the recognition and acknowledgement of “the lived position” of the subject.

As Anita Avramides suggests:

“We do better to reject the philosophical starting point and begin with the position we should maintain is ours: the position that distances itself from conceptual solipsism and

asserts a generality for our mental concepts. In other words, instead of beginning with a philosophical position about our own case, we should assert the lived position of true generality that we all operate with every day. The [conceptual problem of other minds] after all, seemed to arise because we could not see a way of moving from our philosophical starting point to the position of conceptual generality. We avoid this problem if we assume a different starting point” (Avramides 2001: 229).

As I see it, the rejection of the two dogmas enables us to take an important step in overcoming the conceptual problem of other minds, and gives us an opportunity to begin with a new starting point, one that returns us to the lived position of the subject. The lived position of the subject does not involve being conscious of itself *qua* subject— from the inside, intransitively, as an agent, or sans error— but instead as a subject among subjects, a person among persons. From my understanding of the later Wittgenstein, it seems to me he would have admired such a position.

In the next chapter, I consider whether self-ascriptions are IEM, independently of whether the conditional dependence of an account of IEM upon the subject/object distinction. I offer reasons and evidence to think that self-ascriptions are not IEM. In Chapter 5 and 6, I return to an elucidation of what I call, following Avramides, ‘the lived position of the subject’ and in Chapter 7, I discuss the resolution of the conceptual problem of other minds in more detail.

Chapter 4

Subject to Error: Conceptual Arguments and Empirical Evidence to Deny Immunity to Error Through Misidentification

In the last chapter, I argued that we cannot make sense of self-ascriptions being IEM independently of whether such self-ascriptions involve awareness of oneself *qua* subject. I then argued that there is no real principled distinction between awareness of oneself *qua* subject and awareness of oneself *qua* object, because this distinction is usually taken to rest on distinctions that cannot be defended, except if we accept rather unclear notions or are willing to defend positions circularly. The goal of this chapter is to consider, apart from the conditional claim that *if self-ascriptions are IEM, then such self-ascriptions depend upon SC_s*, whether self-ascriptions are IEM in general. I offer two arguments that self-ascriptions are not IEM, one conceptual and one empirical. The conceptual argument is that once it becomes clear that self-ascription is distinct from self-reference, that self-ascriptions depend upon I-thoughts, and that I-thoughts depend upon fundamental grounds of difference from other objects, then it is further clear that self-ascriptions cannot be IEM. The empirical argument is that there are cases of mental self-ascriptions— reports of thought insertion— and bodily self-ascriptions— reports of anarchic hand— that are evidently subject to error. Insofar as these are genuinely self-ascriptions and are genuinely subject to error, then they provide counterexamples to the claim that self-ascriptions are IEM. Thus, self-ascription is not IEM. The role that this plays in the argument for *the Symmetry Thesis* is a defense of the premise that self-ascription depends upon self-identification.

1. Self-ascriptions are not IEM

As I said above, acceptance of the claim that self-ascriptions are IEM is the default position— even for Strawson and Evans— so it would be an accomplishment to make room for the idea that self-ascriptions are subject to error. I think I have made progress on that front by arguing that self-ascription is ascription of experiential predicates to a subject that is an object. However, one might still think that self-ascription, even if it is ascription to an individual particular, is still IEM. One might think that we could argue that self-ascription is IEM independent of the subject/object distinction.

I do not wish to deny that there is a form of IEM that is rather plausible. The plausible form is the form that accompanies self-reference with the first person singular pronoun “I.” When one uses the first personal singular pronoun “I,” one is referring to the speaker of the sentence. The action of carrying out the speech act determines that the producer of the speech act is its reference. In this case, when I say, “I am seeing a sheep,” because of the token-reflexive self-reference rule that governs I-use, ‘I’ refers to the speaker at that time and space. (There are arguments that ‘I’ does not involve guaranteed reference (DeGaynesford (2003), however, and so given sufficient argument and evidence, I would be willing to give up this commitment.). But, even if it turned out that ‘I’ did not have guaranteed reference, given that I have rejected the thought-language principle, this would not have import for the issue of whether I-thoughts have guaranteed reference.)

P. F. Strawson (1994, p. 210) also writes, “it is a rule of [the ordinary practice of personal reference by the use of personal pronouns] that the first person pronoun refers, on each occasion of its use, to whoever uses it.” However, Strawson also suggests that

while the link with criteria of self-reference is not in practice severed, nevertheless, the links between I-thoughts and the empirical criteria upon which I-thoughts rest could be severed. So, I admit that “I” does involve a special type of reference to the individual using it, but this should not be taken to imply either (1) there is a distinction between I-as-subject and I-as-object uses or (2) there is a distinction between two kinds of self-ascription, one subjective and one objective. As Strawson (1994) points out, the linguistic expressions of “I” in self-reference are only thinly immune relative to the first person singular pronoun, but not necessarily relative to the contents of the self-ascriptions. For example, if I think “I am seeing a sheep,” I might have misidentified the sheep, that I am seeing, and (if what I am arguing here is correct), the individual that is the subject of the seeing. Simply because the criteria in practice are never severed between ‘I’ and the speaker— because of the token-reflexive self-reference rule— that does not mean the contents in the subject place and the predicate place are never severed from the empirical conditions that make them possible.

The conceptual argument is that the phenomenon (if even there is such a phenomenon) of immunity to error through misidentification (IEM) only applies to self-reference. Once we get clear about what type of generality is involved in self-ascriptions, on Strawson’s and Evans’ account, then we have to be committed to a form of identification being involved in self-ascriptions. (Of course, Evans argued that bodily self-ascriptions were IEM, but I will argue that he should not have, given his other commitments.) The empirical argument is that in the case of self-ascription, both for ascription of psychological and non-psychological states, there are cases in which awareness of the subject is subject to error.

2. The Conceptual Argument Against IEM

The conceptual argument is that there is a real distinction between self-reference on the one hand, and self-ascription, on the other. I offer two reasons for this: (1) There are self-ascriptions that are non-linguistic in the sense that they do not involve the linguistic expression of those self-ascriptions; (2) If self-ascription were a form of self-reference, then we would run into the paradox of self-reference. Self-ascription depends upon I-thoughts (Bermudez 1998). While we do occasionally posit I-thoughts on analogy with I-sentences, this is not always the case, namely we posit I-thoughts to describe, predict and explain behavior even when we have no need to appeal to self-reference, as I discussed in Chapter 2. I-thoughts do not depend upon self-reference, but instead, I-thoughts depend upon grasping *the δ conception* for oneself as an object. Below, I defend the idea that I-thoughts are dependent upon various empirical criteria. This leads to the conclusion that, therefore, self-ascription cannot be IEM. How does self-reference differ from self-ascription?

Self-reference is the phenomenon of using the first person singular pronoun in sentences of a language, for instance, “I am walking a dog.” However, self-ascription is the ability to ascribe experiential predicates to oneself. While the latter, no doubt, often gets expressed through the former, they are nevertheless distinct phenomena. One might think that the only way to account for self-ascriptions would be to focus on the linguistic expression of self-ascriptions. But, there are several non-linguistic phenomena that require us to say that self-ascriptions are occurring. To take a quite familiar case, one may think “I am seeing a sheep”, “I am throwing a tin can at my sheep”, or “I am angry that my sheep ate my trampoline” and not express these things in action, whether

linguistic or non-linguistic. An infant may think “I am seeing a dog” and not yet have the linguistic resources to say it. As Kant argued, infants self-ascribe prior to being able to say ‘I’. Also, somewhat more controversially, there are social animals that self-ascribe even though they are not capable of saying ‘I.’

These sorts of phenomena require that we account for self-ascription independently of linguistic self-reference. To suppose that self-ascription can only be understood in terms of linguistic self-ascription would suppose what Bermúdez has called the ‘thought-language principle’: “The only way to analyze the capacity to think a particular range of thoughts is by analyzing the capacity for the canonical linguistic expression of those thoughts” (1998, p. 24). I follow Bermúdez in rejecting the thought language principle. As will be shown below, however, I differ from Bermúdez in the kind of analysis of the capacity to think I-thoughts.

One might, however, to postulate that I-thoughts, while they are not underpinned by linguistic self-reference, are nevertheless to be understood in terms of a type of mental self-reference (Brook (2004)). However, many have puzzled by what this special form of mental self-reference could be or if it even exists. Apart from this *prima facie* puzzling nature of the notion of mental self-reference, such a view runs into an explanatory paradox, if one assumes a form of mental self-reference is operable in self-ascription. As Bermúdez (1998, p. 24) has argued, this leads us to a paradox of self-reference. In order to explain how one is capable of learning how to use the first person singular pronoun, one must have an independent account of how one acquires the ability for self-ascription. That account cannot presuppose the ability to use ‘I.’ I outlined the paradox of self-consciousness in Chapter 1.

As I mentioned in Chapter 1, while I agree with Bermúdez that the rejection of the thought-language principle is reasonable in the case of self-ascription, I do not agree with his solution. When we are theorizing about self-ascription and what makes it possible, although we are appealing to I-thoughts to account for self-ascription, this is a placeholder for a functional account of the I-think, not a commitment to the exact same sense and reference of ‘I.’ I have already said that I do not wish to be committed to either conceptual or non-conceptual content of I-thoughts in order to account for self-ascription. However, I am still committed to providing an account of I-thoughts. So, what type of account am I suggesting we use instead? So far, I have only said that self-ascription depends upon empirical conditions. But, I will turn to the account of these empirical conditions and elucidate what they imply about whether self-ascriptions are IEM.

I want to suggest that Evans’ notion of *the δ conception* for an object can help us to understand the type of self-identification that is involved in self-ascription. *The δ conception* involves the identification of the subject. However, it is not incompatible with the phenomenology of seeming to be identification-free. The claim is that although it seems to us that self-ascription is identification-free, there are reasons to think that self-ascription may involve self-identification. As I will argue in Chs. 5 and 6, *the δ conception* of “I” embeds an identification of a subject with respect to the type of individual that one is, therefore, self-ascription requires self-identification as a person. As I discussed in chapter 2, self-ascription depends upon a variety of empirical criteria, which are taken up in *the δ conception*:

The first condition is that the creature has a sensibility (Strawson 1959).

The second condition is that the creature locates itself in space and time (Strawson 1959).

The third condition is that the creature be able to orient itself by tracing out a line in space and time (Strawson 1959).

The fourth condition is that it be capable of distinguishing between appearance and reality (Russell 1994, p. 134).

The fifth condition is that it know-which embodied object it is (Cassam 1994, p. 108).

In Chapters 5 and 6, I will argue that self-ascription must meet a sixth condition, namely the condition that the subject is able to recognize and acknowledge other individuals like itself, or as a person among persons.

It might seem strange that I am suggesting that Evans's view of self-ascription could be based on *the δ conception* that is not IEM, considering that Evans himself claims that self-ascriptions (mental and bodily) are IEM. I believe this actually highlights a deep conflict in Evans' account. As Brinck (1998, §64) argues, "Clearly, if I-thoughts necessarily involved the performance of an explicit identification of the sort $[I=\delta_i]$, they would always be open to error through misidentification, since there is no guarantee that one knows that one is a particular person P, and thereby instantiates a certain description. Actually, it seems that I-thoughts that rely on the proposition $[I=\delta_i]$ are identification-dependent, according to Evans' own definition. This means that I-thoughts could, by definition, not be IEM" (1998, §34). According to Brinck's interpretation, if we focus on Evans' account of *the δ conception*, then insofar as self-ascriptions require a type of I-thoughts, then they cannot *by definition*, be IEM. I agree with Brinck's account.

My suggestion is then that we should tip Evans' account towards the model of self-ascription that rejects IEM. On this account, then, self-ascription would be maximally reliable in normal cases, but not absolutely immune to error. Maximal reliability means that there are two contributions to the self-ascription of experiences:

- (1) the causal relation between the experiential content and the self-ascription of that experience (which is taken to be a matter of causal-information links between, for instance, one's perception of a sheep and one's self-ascription of that perception of a sheep);
- (2) the criterial relation between the self-ascription and the experiential basis or ground for that self-ascription (which is taken to be a matter of grasping *the δ conception* of oneself as a subject of that experience).

It should be obvious from this account that this would require arguing for an empirical account of the kind of identification that might be involved. One way to do that is to consider two cases that put pressure on the claim that self-ascriptions are IEM. My argument is then that if we can discover what is missing in these cases, then we might find a plausible way to say what type of identification is involved in *the δ conception*. So, in this way, I have provided the conceptual room for the claim that a Strawsonian account of self-ascription could be dependent upon an account of self-identification. So, as I discussed at the end of Chapter 3, while the first person singular pronoun "I" is IEM provides a default status assumption that our self-ascriptions, whether they are linguistic or non-linguistic, the latter may depend upon identification, however, and the employment of I-thoughts are occasionally severed from the empirical criteria specified

above. In the next section, I will provide two examples of this, one of mental self-ascription and one of bodily self-ascription.

3. The Empirical Argument Against IEM

To make the empirical argument that self-ascription is not IEM, I need to discuss what is called ‘mental self-ascription’ and ‘bodily self-ascription.’ I do not subscribe to the idea that there is a plausible way to make out a distinction between mental self-ascription or bodily self-ascription. The idea that one is self-ascribing a state to “a mind,” on the one hand, or to “a body,” on the other, is just such a queer idea. To assume that these notions are clear enough would overturn the fact that it is an open question what the ground or basis for I-thoughts are. Further, assuming that we self-ascribe some predicates to minds— mental predicates— and some predicates to bodies— bodily predicates— merely confuses the discussion. There are different predicates, but these are not differences that are exclusive or exhaustive. I will say there are self-ascriptions of psychological states/processes and non-psychological states/processes.

There are cases in which the self-ascription of both psychological states and non-psychological states are subject to error through misidentification. The example of error in the case of psychological states is a well-documented symptom of what is usually considered a disorder of self-consciousness: schizophrenia. I consider the counterexample of individuals with schizophrenia that experience thought insertion. This is a case in which one misidentifies the subject of an experience, and so it qualifies as a type of misidentification because one thinks of oneself that one is not the subject of that experience, when one is the subject of that experience. An example of error in the case of non-psychological states is anarchic hand syndrome. I consider anarchic hand

syndrome a counterexample to IEM, because one misidentifies the subject of an experience. It qualifies as a type of misidentification because one thinks of oneself that one is not the subject of the experience when one is the subject of that experience.

In order to ensure that I am not begging the question that inserted thoughts and anarchic hand are genuine counterexamples, I will review a mainstream description of IEM. Following up on the previous chapter, Shoegenstein suggests that to say that a thought ‘a is Φ ’ is subject to error through misidentification relative to the first person “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” (Shoemaker, 1968, pp. 7–8). According to Joel Smith (2006), a recent proponent of IEM as a mark of self-consciousness, self-ascriptions are IEM relative to certain bases or grounds: “Through introspection it is impossible for me to discover that *someone* resents their mother, and mistakenly think that is *I* who resent my mother. Through talking to one’s analyst, on the other hand, it is quite possible for me to discover that *someone* resents their mother, and mistakenly think that it is me” (Smith, 2006, p. 275). According to Smith, Evans (1982) extends the notion of IEM by discussing the idea of an identification component that is implicit in Shoemaker’s discussion of ‘a is Φ ’. “If a singular proposition is identification-dependent, the thinker knows which object it is about in virtue of an identification. If, on the other hand, a thought does not rest on an identification component, then it is identification-free” (Smith, 2006, p. 275).

According to Smith’s account the common core of IEM given by Wittgenstein, Shoemaker and Evans is as follows: “The proposition ‘a is *F*’ is immune to error relative

to grounds G iff every possible judgement that a is F (based on grounds G) either: (1) does not rest on two beliefs of the form ' b is F ' and ' $a = b$ ' or (2) rests on beliefs of the form ' b is F ' and ' $a = b$ ', but ' $a = b$ ' is not rationally doubtable." (Smith, 2006, p. 276). (I will put aside two issues here. First, I put aside the argument within which Smith is presenting this construal of IEM, because I am not concerned with whether Shoegenstein or Wright's (1998) account of IEM properly accounts for intuitions about self-consciousness (and a consideration of McDowell's (1998a) objections to Wright's account). Second, I put aside the question of whether Smith's account of identifications being 'rationally doubtable' is circular in the way discussed in Chapter 3.).

I will take the liberty of labeling the two forms of IEM that are discussed by Smith here. I will call the first form of IEM 'trivial IEM' because self-ascriptions are immune to error through misidentification in this sense because they do not depend upon identification at all, they are what is called identification-free. I will call the second form of IEM 'non-trivial IEM' because self-ascriptions are immune to error through misidentification in this sense because although they depend upon an identification, that identification is never in error or never fails. I now turn to two counterexamples to this mainstream account of IEM.

The first counterexample focuses on patients with the psychological disorder schizophrenia often report 'hearing voices'. Psychologists and philosophers have called this phenomenon 'inserted thoughts' because the subject often denies that she is the source of those thoughts. A patient might say, "Thoughts come into my head like 'Kill God'. It's just like my mind working, but it isn't. They come from this chap, Chris. They're his thoughts" (Frith, 1992, p. 66). According to Campbell (1999a), the

phenomenon of inserted thoughts discussed here counts as a genuine counterexample to IEM because although the patient is right about which thoughts they are, the patient is wrong about whose thoughts they are. And, the patient is wrong about whose thoughts they are because as Frith (1992) interprets the phenomenon, “This experience implies that we have some way of recognizing our own thoughts. It is as if each thought has a label on it saying ‘mine’. If this labeling process goes wrong, then the thought would be perceived as alien” (Frith, 1992, p. 80).

According to Campbell, the phenomenon of inserted thoughts is best understood in terms of a lack of a sense of agency. Whether or not a thought is experienced as one’s own depends upon whether there is a match between the efference copy of intended thought and the introspection of that thought. So, if we consider the self-ascription of thoughts to be a motor process then, the inserted thoughts phenomenon can be understood in terms of “something going wrong with the monitoring of efferent copies of instructions to think thoughts, so that the subject finds himself lacking in any sense of being the agent of a thought of which he has immediate introspective knowledge” (Campbell, 1999a, p. 615). This does not commit Campbell to the idea that such efferent copies are conscious, that all conscious thoughts are explicitly monitored to ensure that they are one’s own, or that conscious thought is not an executive process. Instead, the account offered here of inserted thoughts is of a mismatch between the non-conscious efferent copy of executed motor intentions to think and the introspective awareness of one’s thoughts.

Why does the phenomenon of inserted thoughts count as a form of self-ascription on my account? According to the functionalist account of self-ascription that I have offered in the dissertation, self-ascriptions are understood in terms of I-thoughts. One

might object that in the case of inserted thoughts, the kinds of self-ascription involved do not meet the functional desideratum. The functional desideratum suggests that the structure has to account for the existential import that self-ascription has for experience in general. And, since the existential import that inserted thoughts is different than the existential import that normal non-inserted thoughts have, then this cannot be a form of self-ascription. However, this presupposes that there is something so abnormal about inserted thoughts that it does not count as a genuine experience by the subject. It may have a different existential import, but it doesn't follow from that fact that it has no existential import.

So, why does the phenomenon of inserted thoughts count as a genuine counterexample to IEM? According to the analysis of the phenomenon presented here, inserted thoughts count as a form of self-ascription and do rest on an identification component that can be in error or fail. To take a simple example, a patient that experiences thought insertion thinks, "I am thinking about Meteora" but then ascribes that thought to Hillary Clinton. The subject uses "I" in thought, but thinks that thought is being had by Hillary Clinton. However, Hillary Clinton is not self-ascribing "I am thinking about Meteora" to herself. So, the subject entertaining the self-ascription, however, has failed to identify the subject of the self-ascription. The error is an error of identification, also, not one of the predicate of the self-ascription, for there is not a question of the subject having a thought about Delphi rather than Meteora, or having a perception of Meteora rather than a thought about Meteora. Notice that, instead, there is a question of who is the subject of that thought.

I now consider an objection to this analysis presented by Shaun Gallagher. Gallagher (2004) describes the phenomenon of thought insertion as an experience where there is a sense of ownership of the thought, for example, “I am thinking about Meteora” is ascribed to oneself as one’s own, however there is a lack of a sense of agency of the thought, since it is ascribed to oneself as arising from another source. However, as Gallagher is right to point out, if it were only a problem of the sense of agency, then the phenomenon of inserted thoughts would be assimilated to experience of unbidden thoughts, like “I left the coffee pot on!!” Gallagher contrasts two possible interpretations of the misascription involved in inserted thoughts. The first is that the cause of the misascription is “based on inferences made at the higher-order level of attributive consciousness” to the effect that someone or something else is the cause of one’s thoughts; the second is that “some neurological component responsible for the differentiation between self and other is disrupted, as a result, some sense of alterity is already implicit in the first-order experience” (Gallagher, 2004). Gallagher presents evidence for the latter account.

While a proper account of the phenomenon is necessary for the discussion, in order to meet Gallagher’s objection it suffices to say that there is a middle ground between these two options that I (following Campbell) wish to tread. First, in order for inserted thoughts to count as a genuine counterexample to IEM, it is not necessary to argue that there is a phenomenon of identificatory inferences made in introspecting one’s experiences. There may be an identification component in self-ascription even if it is not conscious, as Campbell points out. Second, it is not necessary to be committed to the kind of phenomenology of everyday experience that Gallagher suggests is present when

he says that there is some sense of alterity already implicit in the first-order experience. If there were such a phenomenon, then the Symmetry Thesis could be argued directly through phenomenology, but self-ascription is experienced as an ascription to oneself as oneself in normal cases, not as an ascription to an implicit other. Third, although Smith (2006) reads Gallagher's critique of Frith and Campbell's account of thought insertion as providing an objection to thought insertion being a genuine counterexample to IEM, it is Gallagher (2000) that provides that argument. I now turn to that discussion.

It is necessary to stress that according to Gallagher's account of IEM, self-ascriptions are IEM because there is no identification component involved, so Gallagher is committed to trivial IEM: "In other words, we are immune to error in this regard, not because we are so good, or so infallible at judging who we are, but because this kind of self-awareness doesn't involve a judgment at all. We are not in a position to get it wrong, or to get it right; we are simply, in principle or by default, always correct on this score" (Gallagher, 2000). According to Gallagher, however, the phenomenon of thought insertion doesn't count as a counterexample to self-ascriptions being IEM. He says,

"There is good reason to believe that the immunity principle remains intact even in cases of inserted thoughts. When the schizophrenic reports that certain thoughts are not his thoughts, that someone else is generating these thoughts, he is also indicating that these thoughts are being manifested, not "over there" in someone else's head, but within *his own* stream of consciousness, a stream of consciousness for which he claims ownership. In other words, his complaint with regard to inserted thoughts is not that he is suddenly telepathic, but that someone else has invaded his

own mind. The sense of ownership is still intact, despite the feeling that he is no longer the agentive cause of the thoughts. For that reason the schizophrenic should provide a positive answer to what he might rightly regard as a nonsensical question: are you sure that *you* are the one who is experiencing these thoughts? After all, this is precisely his complaint. *He* is experiencing thoughts that seem to be generated by others. His phenomenology is this: he has a sense of ownership for the stream of consciousness which is impossible to misidentify (and is, in fact, in no need of identification, since it is his own), but into which are inserted thoughts for which he has no sense of agency. His judgment that it is *he* who is being subjected to these thoughts is immune to error through misidentification, even if he is completely wrong about who is causing his thoughts. In the latter case, with respect to agency, he is in a position to make only statements in which he uses the first person pronoun as object— and in such cases the immunity principle is not at stake, and therefore cannot be violated” (Gallagher, 2002, my underlined emphasis).

So, according to Gallagher, the immunity principle is not violated, because IEM is based on the sense of ownership rather than the sense of agency involved in self-ascription. Precisely because in the case of inserted thoughts, patients say that they are the owners of the thoughts, then there is no failure of identification, because there is no need of identification. However, as I emphasized in the quote, Gallagher is here mixing up two things that need to be kept separate: the experience of the inserted thought and the self-ascription of the inserted thought. I would agree that in the report of the inserted thought,

the patient does experience that she is the owner of the thought, for otherwise, it would be unusual for him to be able to report on that thought. However, it is not a matter solely of the phenomenology of the *experience* of the thought, but also of the self-ascription of the thought to oneself. Although he is experiencing thoughts that seem to be generated by others he is *self-ascribing* that thought to an other, meaning there is a failure of the identification component of the I-thought involved in the self-ascription.

There is more to be said about whether Campbell or Gallagher are correct about inserted thoughts being a genuine counterexample, but, I think that inserted thoughts counts as a genuine counterexample, and Gallagher's objection here is not sufficient to defend the immunity principle. I now turn to another structurally similar debate between Anthony Marcel and Christopher Peacocke about whether anarchic hand syndrome counts a counterexample to self-ascriptions of action based on proprioception are IEM. It has been argued that anarchic hand syndrome provides a counterexample to IEM (Marcel (2003, p. 81)). Anarchic hand syndrome is a neurological disorder in which it seems that one's hands (or other body parts) take on an agency all their own. One's hand will take things from plates, unbutton one's shirt, or grab at other's body parts. In this case, one reports that one is not performing the action, that some other subject is the agent of the action, when in fact, one is the agent of the action. In this case, it might be argued that anarchic hand syndrome provides a counterexample, because one thinks of oneself that one is not the subject of the action, but one actually is the subject.

What is the mainstream account of anarchic hand syndrome and why does it count as a form of self-ascription? To discuss all the details of Marcel's (2003) extremely rich account of the sense of agency would take us too far afield. Instead, I will focus on what

Marcel says about anarchic hand disorder as a counterexample to IEM. According to Marcel, anarchic hand disorder counts first as an example of the “disownership of action” (2003, p. 76). Anarchic hand is a unimanual phenomenon. Marcel describes anarchic hand as having two components: an experiential aspect and a component of action control. For subjects with anarchic hand, the experiential aspect involves describing the hand as having a will of its own. Patients also often attempt to control the anarchic hand with the other hand, or sit on the anarchic hand, or reprimand the hand as if it were a child acting badly. However, according to Marcel, patients do not report that the anarchic hand is felt as “other” in the sense that a sense of ownership is missing, but instead that the anarchic hand phenomenon is distinctively a failure of the awareness of agency.

Marcel reports one patient's discussion of the lack of the sense of his own intention or urge or effort in anarchic action, “of course I know that I am doing it. It just doesn't feel like me” (2003, p. 79). According to Marcel, for this reason, anarchic hand counts as a failure of misidentification. He suggests that although Evans argued that if knowledge of one's action is via proprioception then “one cannot be mistaken as to whose body one is informed about (Evans, 1982). However, from the point of view of psychology, it is far from obvious that proprioceptive self-knowledge is error-free or that non-observational knowledge is necessarily true; both can be seen as empirical matters” (2003, p. 54). As I argued in Chapter 3, I agree with Marcel that it should be considered an empirical matter whether all self-ascriptions are IEM. Marcel also agrees with me about the undue dogmatic status of IEM when he writes that the claim that the self-ascription of actions are always IEM “has been raised to axiomatic status by some

authors (Shoemaker, 1968; Evans, 1982)” (Marcel, 2003, p. 80). Marcel describes the thesis and anarchic hand as a counterexample in the following:

“The thesis is essentially that any knowledge about properties that is gained in certain ways cannot be known to apply to an individual without *ipso facto* knowing that the properties apply to oneself. While for Shoemaker this holds only for introspectively gained knowledge about mental states, Evans argued that there are ways of acquiring knowledge about our physical properties that are immune to error in the same way, one being through internal proprioception. Since patients’ awareness of the anarchic action need not be via external visual or haptic perception, it would seem that Anarchic Hand disproves this thesis” (2003, p. 81).

So, Marcel marshals evidence for Anarchic Hand being a genuine counterexample to the claim that all bodily self-ascriptions are IEM. Christopher Peacocke (2003) responds to Marcel’s discussion in an appendix.

Peacocke’s first consideration is a matter of describing the phenomenon. According to the account presented by Evans, for a property F, it doesn’t make sense to think “Someone is F, but am I the person that is F?” Peacocke follows Hornsby (1980) in distinguishing between the property of moving one’s own hand and the property of being someone whose hand moves. First, he argues that we should not describe anarchic hand as inspiring the question “someone is moving my hand but is it I that is moving my hand?” because in order for the first conjunct to count as knowledge, it would have to be true that someone was moving one’s hand. However, it is not true that someone is moving one’s hand,

because there is no agent involved in anarchic hand, so no one is moving one's hand. We'll call this 'Peacocke's knowledge objection.' Second, he argues if we describe anarchic hand as inspiring the question "there is an individual someone whose hand moves, but am I the person that is that someone whose hand moves?" then there is no counterexample to IEM, because the patient with anarchic hand does report that he is the someone whose hand moves based or grounded in proprioception. We'll call this 'Peacocke's ownership objection.'

Peacocke admits that Marcel is correct that "the subject of Anarchic Hand can know by proprioception that there is an action of moving an arm, and can wonder whether it's his own action" (2003, p. 109). However, Peacocke argues that Evans' formulation "framed in terms of the question 'Someone is F, but am I the person who is F?' concerns specifically properties of subjects, persons, rather than actions characterized without reference to persons" (2003, p. 109).

According to Peacocke, Evans' account of IEM suggests that there is no gap in the ways of coming to know that a person's mental and bodily properties are instantiated and knowing that those properties are instantiated by oneself.

Peacocke argues that Anarchic Hand does not open that gap because the person with anarchic hand knows through proprioception that there is an individual someone whose hand moves and that someone is oneself. We'll call this 'Peacocke's Persons consideration.' I now turn to my replies to Peacocke's knowledge objection, his ownership objection and his Persons consideration.

Peacocke's knowledge objection is too heavy-handed in its use of the ideas of 'agent' and 'knowledge.' First, with regard to 'agent,' people with

Anarchic Hand do experience the actions as actions of some individual even though they are not actions of anyone *per se*. So, the latter fact cannot be used to rule out the reports of patients on their own phenomenology. Such actions may have no agent, but since patients interact with their hand, often demonizing it or considering it to be its own agent, then that suffices to say that experience they hand as an agent. Second, with regard to ‘knowledge,’ simply because it is not true that someone is moving the patient’s hand, that does not mean that the way of gaining information is not a potential way of knowing about oneself. ‘Someone’ is used as an indefinite specific, but, according to Peacocke’s interpretation specificity implies personhood. But, this assumption need not be sacrosanct. It is not necessary to assume that ‘someone’ implies a full-blooded person rather than merely an agent of the action.

Peacocke’s ownership objection involves a description of anarchic hand that assimilates that phenomenon to Utilization Behavior. In Utilization Behavior, patients’ actions are non-intended actions that are driven by the environment. As Marcel describes, “if on entering the bedroom the bed is visible they will undress and go to bed, or if a matchbox is within reach they will light matches and if a candle is present light the candle” (2003, p. 77). In this case, the subject reports that there is ownership of action, but there is no concern about the action as arising from an agent. So, the description “there is an individual someone whose hand moves” leaves out the pivotal experience of concern about the individual someone whose hand moves.

Peacocke's Persons consideration is important and relevant for my discussion in later chapters because it presents an insight that I think is important for both Strawson's and Evans' account of self-ascription of experiences. However, as I discussed in earlier chapters, there are two ways to approach an analysis of self-ascription, the subject principle and the predicate principle. Although, Peacocke is correct that Evans' focuses on the Person predicates that might be applied to an individual in self-ascription, Peacocke begs the question against Marcel about whether Person predicates are always IEM, because Marcel does use the term 'individual' when characterizing the account IEM, which suggests that he is aware of the Persons consideration. Further, subjects with alien hand often describe the experience of the agent that is controlling the action of the hand with person predicates, so it is unclear why Peacocke's suggested restriction should count against Marcel's claim that anarchic hand is a genuine counterexample.

Both these pairs of counterexamples to mental self-ascription and bodily self-ascription being IEM at least put into question the idea that all self-ascriptions are IEM. This is all in the service of the claim that I started this section with, namely, that IEM is not a mark of self-ascription. Insofar as there was a challenge to the idea that the generalizing of the subject required identification of the subject, that challenge has been overturned. I am now free to assert the claim that I am arguing for in this chapter, namely that there are conceptual and empirical reasons to think that self-ascriptions depend upon self-identifications. But, what should philosophers of self-consciousness do without the idea that self-ascription is IEM? In the last section, I spell out what kind of

approach to self-ascription is viable, namely one that doesn't commit to the two implausible dogmas, one which I discussed in Chapter 3 and one which I discussed in Chapter 4. Before I do, however, I want to address an objection that Shoemaker might lodge against my account of self-ascription.

According to Shoemaker, the reason that we need to appeal to the idea that self-ascription is IEM is to avoid an infinite regress. In everyday life, we do in fact, on occasion identify ourselves via contingent properties. I might do so with the following: I was born in a factory town. I like apple pie and baseball. I have ridden a bucking horse bareback. But, not all of my thoughts about myself can be via contingent properties, but, instead, some thoughts about myself must be IEM, because there is no such identification involved. However, all that is required is that there are thoughts about myself which are necessary, but which do not enter into my introspection of my contingent properties. In this case, I hypothesize that I-thoughts are based upon necessary empirical criteria, but while these do not enter into the contingent descriptions I make of myself, they are necessary in order to make I-thoughts possible.

4. Self-Ascription Without the Dogmas

As I argued in this chapter and the last, a proper account of self-ascription cannot begin with extreme claims like that self-consciousness possesses intrinsic properties of being semantically or epistemologically peculiar. There is little evidence that IEM has this kind of axiomatic status. Even if there were some reason and evidence for this claim, it is not the best way to investigate a phenomenon to appeal to a state or process, in this case, thoughts, which are *in principle* immune to error. In rejecting that self-ascription is IEM, what I take myself to be suggesting is that we allow for the possibility that there is a

form of self-identification that is implicit in our self-ascriptions. It is implicit in that we are not aware of it. While it may not seem or appear present in our own phenomenology, because for instance, when I think, “I am seeing a sheep,” I am not aware of any identification of myself as myself, which should not be the final word. Not all identification is conscious identification in other cases. For instance, I can identify fire trucks in my sleep, but I am not conscious of them in any way. So, why should we suppose that self-identification must be conscious? Regardless, the idea that self-consciousness is IEM can only be bolstered by a kind of unreflective phenomenology.

Such an approach is not the final arbiter, but instead, a type of reflective phenomenology should be employed. What this means is that we should think about our self-ascription in terms of a more holistic account which brings in resources from outside our occurrent experience of introspection. That means that anthropologists, developmental psychologists, and evolutionary biologists, to name a few, have something to contribute to consideration of how at some point or other identification might have been involved occurrent self-identification. It is just that self-ascription has become so well worn in our mental economy that we now take it for granted. We might do some research into other cultures to see if there are any widespread experiences of inserted thoughts or anarchic hand phenomena. We might consider the early experiences of infants to begin to theorize about whether their experience can be captured in these terms. Similarly, we might theorized about our phylogenetic ancestors’ experience of their thoughts and actions.

What I am suggesting is that it might be wrong to think that there is a special sphere of ascriptions which are IEM, but rather that there are well-worn identifications

that have become so reliable in our mental economy that those identifications rarely produce errors. The takeaway slogan here is the following: *maximally reliability should not be taken to entail absolute immunity*. This position brings with it a beneficial consequence. Without the claim that self-ascription is IEM (even “thinly IEM”), self-ascription is a process which can more easily be understood in naturalistic terms. There will be laws that govern the reliable process, but there will also be occasions where the process breaks down. These should be manifested in disorders and fringe cases, but the breakdowns are nevertheless instructive for the approach of reflective biological phenomenology.

I have argued that assuming that self-ascription involves some type of generalizing of the subject, then it might involve a form of subject identification. I have made room for this possibility by arguing against the claim that self-ascription is IEM. I argued that we cannot make sense of IEM, unless we can articulate a distinction between awareness of oneself *qua* subject and awareness of oneself *qua* object. But, that distinction is a mere philosophers’ myth. I also argued that there are reasons independent of the conditional that if IEM, then subject/object, that IEM is questionable in its own right. In that sense, I have piled on skepticism about IEM in order to make room for the idea that identification is involved in self-ascription, particularly in identifying the subject of self-ascriptions.

In the next chapter, assuming that I have been successful, I consider two principles that might govern how that identification is achieved. On one account, *the predicate principle*, the identification of the subject is made possible by relying on the predicates of self-ascriptions. On another account, *the Persons principle*, the

identification of the subject requires identification of the subject independent of those predicates. I discuss these two principles— the predicate principle and the Persons principle— in detail. I argue that the subject of self-ascription has to identify itself outside the subject of its experiences. This leads to the conclusion that we identify ourselves primarily as persons among persons in a particular form of life.

Chapter 5

Against Cognitive Descriptivism: Self-Ascription, Identification and the Persons Principle

Suppose a subject self-ascribes a perceptual experience. For example, the subject thinks, “I am seeing a sheep” when a sheep appears before her. Setting aside the question of whether it is actually a sheep and whether she is actually entertaining a visual perceptual experience, we might ask, “How does the subject come to identify herself as the subject seeing the sheep?” To answer this question, we need to know what makes self-identification possible. (I will suppose that in order to self-ascribe an experience, some form of self-identification is involved, given that I have argued for this claim in the previous two chapters. For example, in order for an individual to have an experience that is an experience *for one* or experienced as one’s own, one needs to identify oneself (at least minimally) as the subject of that experience. I admit this is controversial, because it is commonly assumed that self-ascriptions are immune to error through misidentification (IEM) relative to the first person singular pronoun. However, this is not a sacrosanct assumption, as I have argued and it has come under fire in the discussion of self-ascription: (Tugendhat (1986, p. 297), Strawson, P. F. (1994); Cassam (1997, p. 61–2); Bermúdez (1998: Chapter 9); Rosenthal (2005; Chapter 13); Meeks (Diss.); Schwitzgebel (forthcoming).) In this paper, I will be contrasting two distinct answers. One presents a Lockean view of self-identification, which proposes that self-identification is achieved through the experiential predicates that are self-ascribed. Another answer presents a Kantian view of self-identification, which claims that self-identification is achieved through the recognition and acknowledgement of oneself as a person among persons.

The Lockean view suggests that the identification of the subject is based upon what passes before the subject's mind in self-ascription, for instance, the perceptions, actions, or emotions that are ascribed to the subject. The Lockean view is guided by what I will call 'the predicate principle of self-ascription.' The predicate principle suggests that self-identification is relative to a psychological sortal that provides a cognitive description of the subject over time, describing the subject as a perceiver, agent, emoter. The predicate principle suggests that the subject of self-ascription is determined by the content of the predicates that are ascribed to the subject. I will interpret the Lockean view as a kind of *piecemeal cognitive descriptivism*, under which the subject is described over time by the various contingent cognitive contents of the self-ascribed predicates. In §1, I outline the Lockean view in more detail by considering José Luis Bermúdez's (1998, p. 229–265) Neo-Lockean account of self-ascription.

In §2, I raise three objections against the Lockean view.

(1) *the ego objection*: I argue that the Lockean view is committed to the existence of a problematic psychological subject. I point out that there is a trend in Lockean accounts of self-identification of drawing an implausible bifurcation between awareness of oneself *qua* subject and awareness of oneself *qua* object. The diagnosis I offer is that the predicate principle that underpins the Lockean view commits a category mistake: it infers from the idea of a set of psychological predicates to the idea of a special psychological subject. By way of contrast, the Kantian view argues that there is no distinct psychological subject from the practical, everyday embodied person.

(2) *the functional unity objection*: I argue that the predicate principle leads to a multiplicity of subjects, which rules out accounting for the functional unity of the subject. The Kantian view accounts for the functional unity of self-ascription by positing the I-think, which provides for the functional unity of the subject.

(3) *the lived embodiment objection*: I argue that there are other modes of identification which are central to self-ascription other than the contents of perception, action and emotion, namely those having to do with being an embodied living creature. The Kantian view argues that the subject of self-ascription is identified by being a person among persons, a human being among other human beings (Strawson (1966, p. 102)).

Once I have provided these objections to the Lockean view, in §3, I present a positive argument for the Kantian view. P. F. Strawson (1959; 1966) and Gareth Evans (1982) provide an account of self-ascription that supports what I will call the ‘Persons Principle’ of self-ascription. Building on an argument by Strawson, I argue that self-identification depends upon registering the centrality of the concept of the person. This provides for a genuinely different view of self-identification than the Lockean view because it does not rely upon a psychological sortal that derives its content from the predicates that are self-ascribed— perceptions, actions and emotions— but instead, from empirical criteria of persons independent of those predicates. I argue that once the Persons Principle is made clear, an account of self-identification comes into view that depends upon what Anita Avramides (2001) has called ‘the lived position’ of the individual, which takes into account how self-identification is dependent upon recognition and acknowledgement of

being a person among persons. (Though I will not discuss their accounts, this appeal to the lived position of the individual is consistent with claims made about the primacy of the personal level subject made by Merleau-Ponty (1981), Donald Davidson (1996) and Alva Noe (2009).) I close by discussing how this Kantian view of self-identification might be used to generate a genuinely new account of how we are aware of other minds, one that contrasts with the theory-theory and the simulation theory.

1. The Target: The Lockean View

The question of this chapter is, “What makes possible the identification of the subject of self-ascription?” For example, when I think “I am seeing a sheep,” how does the subject come to identify herself as the subject seeing the sheep. There are two possible principles that lead to two different types of answers to this question.

The Lockean view suggests that self-identification is based upon a psychological sortal that embeds a variety of cognitive descriptions of the subject. The principle that guides the Lockean view, I will call ‘the predicate principle’: *in order to identify the subject of self-ascription, the subject thinks about herself under a psychological sortal which gets its content from the experiential predicates ascribed, for instance, perceptions, actions and emotions.* The proponent of the predicate principle suggests that the content of the predicates— for example, “x is seeing a sheep,” “x is walking a dog,” “x is elated about her victory”— carry with them all that is necessary to provide for self-identification, and for this reason it is a claim about necessity and sufficiency of cognitive descriptions of a psychological subject to do the work of self-identification. The Lockean view deserves to be called ‘Lockean’ because of Locke’s focus on internal reflection or

inner sense (what we call ‘introspection’ today) as the context for inquiry into self-identity.

The Kantian view suggests that self-identification is based upon contents external to the content of the experiential predicates involved in self-ascriptions— what might be called ‘transcendental empiricist criteria.’ Strawson defends such a position in his famous essay “Persons.” The principle that guides the second answer I will call “the Persons Principle”: *in order to identify the subject of self-ascription, it is necessary to be aware of the personal level subject independently of the experiential predicates ascribed.* The Persons Principle is a claim about a non-negligible condition of self-identification. In particular, the proponent of that principle argues that the content of the experiential predicates are neither necessary nor sufficient to provide for self-identification. The Kantian view deserves to be called ‘Kantian’ because of Kant’s skepticism about introspection— internal reflection or inner sense— being able to provide an account of self-identity. My reading of Kant’s account of the I-think is influenced by Sellars’ (1970) account of the primacy of the practical subject. My goal in the paper is to defend the Kantian view of self-identification against the Lockean view.

One defense of the Lockean view is presented by Bermúdez in The Paradox of Self-Consciousness. He argues that self-ascription involves a form of self-identification. According to Bermúdez, the identification of the subject of self-ascription is achieved through the predicates of self-ascription. Bermúdez (1998, p. 234) argues that since self-ascription is a type of *psychological* self-consciousness, the subject must be a psychological subject. A psychological subject is minimally, according to Bermúdez, a subject that entertains psychological or mental states, such as perceptions, actions and

emotions. According to Bermúdez, the concept of a psychological subject is a theory-cluster concept (2005, p. 6–13), meaning there are no necessary or sufficient conditions that enter into determining which subject is identified. He focuses on the core notion of a psychological subject in the following principle: “Psychological subjects with a perspective on the world are aware of themselves as perceivers, as agents, and as having reactive psychological states” (1998, p. 247). Given that a subject ascribes these contents to itself, the subject of self-ascription is represented by a host of descriptions embedded in the contents of experiences, thereby identifying oneself as a perceiver, an agent, and an emoter. Bermúdez’s account of self-identification is committed to the predicate principle, because the nature of the subject of self-ascription is determined by the predicates that are ascribed to that subject. I will call the Lockean view that Bermúdez is committed to ‘piecemeal cognitive descriptivism.’

The Lockean view of self-identification is regarded as ‘piecemeal’ because it is over a period of time that one comes to be able to identify oneself as the subject of experience. Self-identification does not happen in one fell swoop according to this account, but through multiple occasions of cognitive descriptions employing a myriad of contingent properties. The theory is ‘cognitive’ because it is through contents such as beliefs, thoughts and ideas that one is described, not through any other mental states, such as sensations or perceptions. Self-identification, according to this account, is underwritten by mental states that are thought-like, rather than, for instance, perception-like. The theory is a form of ‘descriptivism’ because the contents that enter into the identification of the subject serve as definite descriptions of the psychological subject. To identify a self, accordingly, is to describe and interpret the subject through in inner

narrative about oneself, rather than to be aware of itself through a non-descriptive mode of awareness. In §3, I address the objection that if descriptions do have an indexical or demonstrative component, then they are not truly descriptions. I argue that if such Lockean descriptions do not have an indexical or demonstrative component, then they cannot refer singly to a subject of experience. The Kantian view, by contrast, suggests that demonstrative reference to persons among persons is central to an account of self-identification.

As Locke points out, the subject of experience is a “thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places; which it does only by that consciousness which is inseparable from thinking, and, as it seems to me, essential to it” (Locke II.27.ii). Locke’s account is then that self-identification— considering oneself as oneself— occurs via conscious thinking about oneself. In this way, Locke is committed to piecemeal cognitive descriptivism, which accounts for the subject as a cognitive representation that is described over time and amasses to a psychological sortal, under which self-identity is possible. Now that the target of the criticism has been made transparent, I now turn to the objections to the Lockean view and predicate principle.

2. Egos, Functional Unity and Lived Embodiment

The Ego Objection

The ego objection argues that proponents of the predicate principle are justified in inferring from the idea of (P)— a subject that self-ascribes psychological predicates— to (Q)— that subject of self-ascription is determined through piecemeal cognitive description by those psychological predicates. The inference leads Lockeans to draw an

implausible bifurcation between awareness of oneself *qua* subject and awareness of oneself *qua* object.

One manifestation of the Lockean view is in Cassam's account of self-consciousness in Self and World. Cassam (1997, p. 2) argues for a materialist conception of self-consciousness, according to which we are aware of ourselves as physical objects among physical objects. He understands self-consciousness to be "bound up with a sense of oneself *qua* subject of experience as a physical object" (1997, p. 30). As was made clear in Chapter 3, and will become clear in §3, I support the goal of arguing for a materialism about self-consciousness. However, what is puzzling is why Cassam should take himself to be committed to self-consciousness being *qua* subject.

Bermúdez also suggests a similar bifurcation between awareness *qua* subject and awareness *qua* object when he writes: "a minimal requirement for something to be a psychological subject is surely that at least some psychological predicates must be true of it." To take our familiar example, then, a minimal requirement for something to be a psychological subject is that the set of psychological predicates 'seeing a sheep,' 'walking a sheep', 'being angry at a sheep' are true of it. Bermúdez suggests that what enters into being a psychological subject is being a perceiver, agent, and subject of emotions. Even though a minimal requirement for being a subject of self-ascription of experiences is having experiential predicates be true of that thing, it does not follow that what makes something the subject of self-ascription are those psychological predicates. But, why should Lockeans such as Cassam and Bermúdez be committed to positing an extra subject *qua* psychological subject, which involves entertaining consciousness, perceptions, actions, emotions, moods, feelings, itches, pains, tickles, etc.?

Bermúdez's calls self-ascription "the explicitly psychological dimension of self-awareness" (1998, p. 229). Through self-ascription, the subject is aware of "the embodied self as a bearer of psychological properties" (1998, p. 229). In both Cassam's and Bermúdez's accounts of self-ascription, therefore, there is a form of representational ascent— i.e., the ascent from being aware of a subject *simpliciter* to being aware of a subject-as such and such— which gives the subject of self-ascription a bi-level structure.

A proponent of the predicate principle might reply that the concern with self-ascription is predominantly with the ascription of psychological predicates. The paradigmatic form of self-consciousness that Bermúdez is concerned with is psychological, namely, the self-ascription of psychological predicates. So, a proponent of the predicate principle might argue that since we are concerned with psychological predicates, then we are only concerned with psychological subjects.

A response to this reply is that the inference that generates the predicate principle commits a category mistake. It does not follow from the fact that self-ascription is sometimes the ascription of psychological (or non-psychological) predicates to a subject, that the nature of that subject is psychological (or non-psychological). As Gilbert Ryle argues, inferences like those leading to the predicate principle "represent the facts of mental life as if they belonged to one logical type or category (or range of types or categories), when they actually belong to another" (Ryle (1949, p. 16)). In this case, it represents the facts of self-identification, namely, what it is that experiences are ascribed to, as if they belong to the category of psychological subjects, when they actually belong to the category of persons. As Strawson argues (echoing the Rylean point): "There would be not question of ascribing one's own states of consciousness or experiences to

anything unless one also ascribed, or were ready and able to ascribe, states of consciousness, or experiences, to other individual entities of the same logical type as that thing to which one ascribes one's own states of consciousness" (1959, p. 100). If one truly wants to provide a materialist account of self-ascription, then one must account for the self-ascription of predicates *in general*, not just psychological predicates. Further, one must leave open the possibility that self-ascription is ascription to a subject that possesses corporeal properties. "I am sitting with my legs crossed" is no less a self-ascription than "I am seeing a sheep." Thus, the predicate principle falls ill to the ego objection.

the functional unity objection

The second objection to the Lockean view and the predicate principle is the functional unity objection. If we start with a constellation of predications, for example, perceptions such as "I am seeing a sheep," actions such as "I am walking the cow" and emotions such as "I am angry at my obnoxious rooster," then the Lockean ends up being committed to the rather implausible idea that self-ascriptions are dependent upon a multiplicity of subjects, absent any ways or means by which these subjects are combined. The theoretical commitment to the predicate principle turns the subject of self-ascription into a ceaselessly multiplying subject: a perceiver, an agent and an emoter. But, in order to make this objection stick, I have to argue that self-ascription must meet the condition that self-ascriptions are functionality unified.

One fundamental desideratum for any account of self-ascription is to provide for the functional unity of experiences that are self-ascribed. Self-ascriptions depend upon I-thoughts (Frege (1956); Strawson (1959); Evans (1982)). It is usually assumed that I-

thoughts have import for perception, action, and emotion in ways that other thoughts do not. For instance, John Perry (1979) argues for what I will call “the functional unity desideratum”: any theory of self-ascription has to be able to account for the existential import that self-ascription has for experience in general. (Kant (1781) makes this point in the Paralogisms when he argues that the I-think is an existential category (B406ff).) It is called the ‘functional unity desideratum,’ because I-thoughts enable one’s experiences to have a functional unity that would not be possible without such thoughts. Any theory of self-ascription has to be able to account for the existential import that self-ascription has for experience in general. In the following, I argue that because the mutual reciprocal dependencies of perception, action and emotion on I-thoughts, the functional unity desideratum holds of any account of self-ascription.

I-thoughts have import for perception in being related to here-thoughts and now-thoughts, which illustrates their basis in empirical criteria (Evans (1982, p. Chapter 7)). For example, if I think “I am seeing a sheep,” then my self-ascription is dependent upon ways of gaining information about the here and now, for instance, that my body is taking up such and such space and that I am situated in such and such a moment. Also, my ways or modes of gaining information through perception are dependent significantly on my I-thoughts; my spatiotemporal position is grounded in my capacity to have I-thoughts. For example, unless I were able to think “I am seeing a sheep,” I could not think “this sheep is here and now.” The ways that our perceptions represent spatiotemporal position independent of us is dependent upon the ways that we represent our own spatiotemporal position. For example, one could not use a map of a space unless one was able to orient

oneself with a “You are here” sign in the space. In addition, one could not locate oneself as ‘here’ unless one could locate oneself in an objective world (Ismael (2007, p. 216)).

I-thoughts also have immediate import for action. For example, if I think “I am going to be eaten by a bear,” then this has immediate import for action in a way that “James is going to be eaten by a bear” may not, because I may not know that I am identical with James (because, for instance if I am an amnesiac, I may forget my name). Also, one’s ability to engage in the self-ascription of action also depends significantly on one’s I-thoughts. We do not say that one is able to self-ascribe one’s actions unless one is also able to affix to one’s actions, “I am doing such and such.” If we could not engage in self-ascription, then our bodily movements would be nothing to us (Kant (1781); McDowell (1994; 2007)). They would seem to us to be alien— someone else’s actions, mental or bodily— or they would be occurrences in one’s body that were inaccessible to our awareness.

I-thoughts also have import for emotion. For example, if I think, “I just found out that my Grandma died” then this has import for my emotional life in a way that “James just found out his Grandma died” does not. I-thoughts have an immediate import to one’s emotional responses, in the sense that one has care and concern for objects of response that bears on one’s existence. In order for emotions to have impact on my perception and action, I need to be able to think “I-thoughts” about my emotions in such a way that it enables those emotions to be pervasive in my experience. For example, if I am to be said to perceive a prey animal and flee from a prey animal, then it is required that think “I am afraid of the prey animal.”

In each of these senses, therefore, the functional desideratum must be met. If self-ascription depends upon I-thoughts and I-thoughts are mutually reciprocally dependent on the contents of perceptions, actions and emotions, then there must be a functional unity of self-ascription that is achieved through I-thoughts. However, the Lockean view and the predicate principle cannot account for functional unity. Suppose we follow the predicate principle to account for three distinct experiences: “I am seeing a sheep”; “I am walking my sheep”; and “I am angry at my sheep”. According to the predicate principle, in order to identify the subject of self-ascription, it is sufficient to consider the content of the experiential predicates ascribed. But, the principle goes against the functional desideratum, for the following reason. Suppose the subject sees her sheep eat her car tire, while walking her sheep, and has the response of anger towards the misbehaving animal. If we want to account for how her diverse mental states were self-ascribed to a single subject of experience, we would have to make explicit how the perception, the action and the emotion were functionally unified. However, the predicate principle leads us to posit three distinct subjects, the visual perceiver, the walker, and the angered individual. Further, the absurdity does not end there, because the predicate principle leads to an endless multiplication of subjects. For instance, if we focus on the range of possible predicates of perception, we might be forced to posit a seer, a hearer, a smeller, or a wine taster. If we focus on the range of possible predicates of action, we might be forced to posit a ball-thrower, a hand wringer, or a Marathon runner. Similarly, if we consider the range of possible predicates of emotions, we might be forced to put forward a griever, a manic, or a melancholic. The issue is that without a substantial basis for the subject of self-identification, the subject multiplies into too many subjects.

The proponent of the predicate principle might simply object that the functional unity of the subject of self-ascription is the product of a more general representation of the subject. For instance, Andrew Brook suggests that we can remedy this problem by appealing to “a representer” as the basic category. The Lockean would then posit that there is a substratum that binds each of these distinct subjects into one subject. (See for instance, Brook’s (1994) Lockean interpretation of Kant’s I-think). Such a representation would serve to connect our perceptions, actions and emotions, but it would also enable self-identification by connecting multiple identifications into a single subject that is represented through piecemeal cognitive descriptions. A proponent of the predicate principle suggests that the functional unity objection minimizes the capacities of organisms to grasp complex representations of ourselves.

The Lockean might argue that have complex representations of ourselves that include being a perceiver, being an agent, being an emoter in a way that does not rule out functional unity. After all, creatures can grasp a representation of an object that has many constituent representations, for instance, the concept of a dairy cow has the property of having utters, producing milk, having four legs, being of a certain massive size. The proponent of the predicate principle argues that there is no particular trouble in the case of a cow, so there should not be in the case of the psychological concept.

My response to this reply is that unlike the constituents of the concept of a cow, the positing of a higher-order content that describes the subject of experiences simply introduces more content, which cannot provide an explanation for functional unity, because it presupposes functional unity. As Susan Hurley (1998, p. 97) has argued, one cannot account for the functional unity of experience by introducing just more contents,

because those contents themselves would need to be functionally unified with the rest. As she argues, without some type of objective contents, “self-consciousness or first-person contents (if they are possible at all) are *just more contents*” to which the problem of functional unity applies. For example, if the subject sees her sheep eat her car tire, while walking her sheep, and has the response of anger towards the misbehaving animal, then the perception, action and emotion need to be functionally unified by an I-thought. However, if one posits a content that over time cognitively describes the subject, then that content also has to be functionally unified with the initial contents. Adding more subjective contents to the mix just requires more contents; what is needed are objective contents independent of the contents of the predicates of self-ascription that provide for combination. The introduction of more contents presupposes that the problem of functional unity is answered, and so cannot be provided as an answer to that problem. As long as one assumes that the predicates do the work of affixing the identification of the subject, then the functional unity objection provides a reason to reject the predicate principle.

the lived embodiment objection

The lived embodiment objection to the Lockean view suggests that there are other types of self-identification that are central to self-ascription other than those that can be achieved via the psychological predicates of perception, action and emotion. The predicate principle does not allow us to consider that self-identification might depend on a form of recognition and acknowledgement of persons in perception. That primitive mode of experience of persons does not enter into the typical predicates that Bermúdez

outlines in perception, action and emotion. The proponent of the predicate principle has to make room for what I call ‘the lived position of the individual.’

Perception, action and emotion all depend fundamentally upon lived embodiment. For example, seeing requires eyes that are functioning, located in such and such a place, and able to be oriented towards objects. Actions such as throwing requires arms, or at least, the synthetic analogue to an arm. Emotion requires having a heart and viscera. But, while these contents do not appear in the content of the predicates ascribed in self-ascription, because those contents are transparent— they are about the world, not oneself— one cannot self-ascribe those predicates of perception, action and emotion, unless one can at least denote that one is embodied. As I mentioned above, this requires that the subject register what Hurley (1998) calls an objective content to one’s I-thoughts independently of one’s experience. But, the content of the predicates themselves do not (and could not) carry that content, because they describe the subject over time via cognitive contents.

If I am perceiving, moving and feeling, then the content is not the body itself but whatever I am perceiving, moving toward and feeling about. Notice that this does not change when these experiences are self-ascribed. For example, when one self-ascribes with “I am seeing a sheep,” the content of the predicate is merely the seeing of the sheep. When one self-ascribes “I am walking the cow,” the content of the predicate is merely the walking of the cow. Similarly, when one self-ascribes “I am angry at my rooster,” the content of the predicate is the anger directed at one’s rooster. None of these actually describe the subject at all; but instead, the experiences of the world independent of the subject. In each of these cases, the content of the predicates themselves are directed

towards the objects of the experiences, not towards the subject of the experiences. So, in order to account for the identification of the subject, one must go outside the content of the predicates and presuppose the lived embodiment of the subject.

One reply to the lived embodiment objection that Lockeans have at their disposal is the idea that some form of sentience is sufficient for self-identification. As I mentioned above, Lockeans bi-level view might posit a substratum of more primitive self-awareness, which then is manifested at a higher level as sapient self-awareness. The idea presented here is that there is a distinction between sentient self-awareness and sapient self-awareness. Sentience involves lived embodiment in a fundamental way because it considers the lived body to be basic to the response to stimuli, which is found in Bermúdez's account of self-consciousness at levels prior to self-ascription: ecological self-awareness, proprioceptive awareness, awareness of one's embodied agency, an experiential point of view. According to Bermúdez's account, there is a real distinction between the distinctive type of lived embodiment that is taken up in self-ascription, because the psychological subject represents "the embodied self as a bearer of psychological properties" (1998, p. 229). The reply suggests that it may be that sapience is dependent upon sentience. For example, when one ascribes perceptions to oneself, one ascribes perceptions of one's own body through interoceptive and exteroceptive senses. When one ascribes actions to oneself, one ascribes a variety of behaviors that are distinctive of our awareness of our actions. When one ascribes emotions to oneself, one ascribes states from a subjective point of view. Since these all contain contents that describe the lived embodiment of the creature, then lived embodiment objection is misplaced.

My response is that if one distinguishes between sentient and sapience to account self-ascription, then this leads to a classic Lockean dilemma. Either one needs to account for sapient self-awareness at the level of material objects, namely the person as a fundamental material particular— which would deny Cassam’s and Bermúdez’s tendency towards representational ascent— or the connection between one’s sapient identification— in this case, identification via the predicates of self-ascription— is difficult to connect with sentient identification. Bermúdez’ account leaves a rent between self-ascription and other more primitive forms of self-consciousness, as has been pointed out (Brook (1999)).

The Kantian view obviates this dilemma. In P. F. Strawson’s famous essay “Persons,” in *Individuals*, he argues that one cannot identify a subject of experience if one can only identify them only as subjects of conscious experience (1959, p. 96). According to the Kantian view, self-identification cannot be made within consciousness itself, but must be done independently of one’s self-ascription of experiences (Cf. McDowell (1996, p. 99–104)). The Kantian view then presupposes that there is a subject *qua* person to which one’s experiences are ascribed.

3. For the Persons Principle: A Kantian Argument

The claim that personhood is prior to a subject within consciousness suggests that if self-ascription is to be grounded in any subject, then self-ascription must be grounded in a subject external to the stream of consciousness. Given that the concept of a person is primitive to our practice of self-ascribing, self-ascription is grounded in the concept of a person. This suggests, therefore, that Strawson was committed to the Persons Principle. However, Strawson’s insight with respect to self-ascription is that one must recognize

how the lived embodiment of the organism, in his terms “the person,” is fundamental to the self-ascription of the most basic of psychological and non-psychological predicates.

Assuming that I have argued successfully that the predicate principle is prone to the above objections, what reasons or evidence exist in favor of the Persons Principle? I have only offered negative support by critiquing the predicate principle with three objections. I have offered no positive support for the Persons principle. In closing, I will paint a picture of the type of view that I think follows from the Persons Principle and is meant as a competitor to piecemeal cognitive descriptivism.

What is distinctive about self-identification governed by the predicate principle is that a subject is cognitively described over time by its predicates. This type of representation is not only insufficient, but the description of the subject is not even necessary. A proponent of the Kantian view might argue that the clearest support for the Persons Principle is the idea that there is available an independent subject— the Person— available that makes possible self-ascription.

Alva Noe (2007; 2009) has recently brought an anti-representational argument to the fore that can be used to make the point. One basic assumption of the Lockean view is that self-identification must employ piecemeal cognitive descriptions to represent the subject of self-ascriptions. Usually, representations such as psychological sortals are posited in the absence of the object. However, the anti-representational argument suggests that if the person as a biological organism is present— as opposed to absent— then there is no reason to posit self-representations via predicates that are necessary for self-identification. There is already available such a biological organism. Namely, the person is conceptually prior to the putative psychological subject that is proposed by the

predicate principle. As Strawson argues, “the concept of a person is logically prior to that of an individual consciousness” (1959, p. 99). Therefore, the representation of the subject is not necessary for subject identification, since the subject as a person is available.

In a different context, something similar was argued by Marjorie Grene (1972), when she suggests that mind and body should be understood through “new categories, primarily in terms of the category of ‘person’ rather than either body or mind.”

According to Marjorie Grene’s view, the mind-body problem can be dispelled through the privileging on persons. In order to dispel the conceptual problem of other minds, Anita Avramides (2001, p. 236) has called the notion that the person is primitive ‘the lived position’ of the biological organism. The Persons principle is further defended by the idea that the identification and the re-identification of organisms can be rather easy.

Consider for example, the phenomenon of joint perceiving. For instance, infants at around the age of 9 months begin to exploit perceptions of themselves and others as persons in order to participate in a dance of gazing and pointing that enables joint engagement with objects. As Hobson (2002, p. 190) argues “there is a mode of feeling perception that is critical for establishing intersubjective relations between people, and it is a kind of perception that establishes a special quality of relatedness between the individual and what is perceived— in most natural circumstances, a person” (190).

According to the account I am offering here, through direct perception, we are aware of ourselves and others as persons. Why is the Kantian view a real alternative to piecemeal cognitive descriptivism? Is the Kantian view susceptible to the three objections lodged against the Lockean view?

There are three important features of the Kantian account that follows from the Persons Principle. The first feature is that self-identification is not achieved via a piecemeal process, but instead, it is accomplished in what Stanley Cavell has called a “single whirl of the organism” (2002, p. 52) or the form of life of the creature, namely the recognition and acknowledgement that a creature is a person among persons. I would argue that a creature comes to self-identify by recognizing and acknowledging that it is person among persons, such that self-identification is not an internal process of a single organism, but is taken up in an external process of the socialization and cultural formation of the subject. The second feature is that self-identification is not achieved via cognitive representations, but instead depends fundamentally upon identification of one’s lived embodiment. While self-identification surely depends upon having perceptions, actions and emotions, bodily self-perception involved in these contents is crucial, rather than the cognitive representations that might be taken up in the experiential predicates. The third feature is that self-identification is not dependent upon descriptions, but instead is a matter of non-descriptive relation to oneself as oneself. I would argue that self-identification is achieved through direct perceptual awareness rather than cognitive descriptions of the subject. I have only provided a picture of the type of view that I think follows from acceptance of the Persons Principle.

Before I close, I would like to show that the Kantian view and the Persons Principle are not vulnerable to the above objections. The Kantian view is not susceptible to the ego objection, because it does not posit contents that represent a psychological subject. As Kant has argued, the I-think is an empty subject, void of content (B406) that does not introduce the distinction between awareness of oneself *qua* subject and

awareness of oneself *qua* object. According to the Kantian view, the I-think that makes possible self-ascription is grounded in direct perceptual experiences of oneself as a person among persons. The Kantian view is not prone to the functional unity objection, because the Persons Principle provides for the functional unity of the subject. The recognition and acknowledgement of oneself as a person among persons determines that one is a functionally unified subject. Also, the Kantian view is not vulnerable to the lived embodiment objection, because according to the Persons Principle, the person is primary, and the best account of the person is understood in terms of its embodiment as a human animal.

I have said very little about how this account can be discussed in the context of the cognitive sciences. The hypothesis on offer here is that self-identification is achieved by the recognition and acknowledgement of oneself as a person among persons. According to this account, relative to perceptions of the type of organism that one is—perceptions that are taken up in relations between subjects—the subject of self-ascription is identified. If this is the case, and self-ascription is dependent upon the identification and re-identification of the lived position of organisms, then several questions can be raised about how exactly we are aware of other subjects.

It is customary to assume that there are two accounts of how we are aware of the minds of other subjects, what is sometimes called ‘mentalizing.’ One account is the theory-theory, which suggests that it is through thinking about other subjects—mainly, through explanation and prediction—that we are aware of other minds. Another account is the simulation theory, which suggests that it is through imagining ourselves in the other subject’s point of view—mainly, through simulating the other’s experiences—that we

are aware of other minds. But, the Persons Principle leads us to the idea that it is the recognition and acknowledgement of ourselves as persons among persons that is central for self-ascription. However, if this were the case, then as Axel Seemann (2008) has argued, it is the perception of persons that is central for mindreading, not thought about or imagination of other minds. This suggests, therefore, that the Kantian view presents a genuinely new theory of mentalizing.

I have contrasted two different views on self-identification: the Lockean view and the Kantian view. I argued that the Lockean view is committed to the predicate principle and is a type of piecemeal cognitive descriptivism. I provided three devastating objections to the Lockean view: the ego objection, the functional unity objection, and the lived embodiment objection. I also provided a brief positive argument for the Kantian view. On the Kantian view, the concept of the person is primitive for self-identification. As Strawson has argued, the concept of a subject of experience cannot be a primitive concept, “at least, cannot exist as a primary concept in terms of which the concept of a person can be explained or analyzed. It can exist only, if at all, as a secondary, non-primitive concept, which itself is to be explained, analyzed, in terms of the concept of a person” (1959, p. 99).

Chapter 6

On the Joint Engagement of Persons: Self-Consciousness, the Symmetry Thesis and Person Perception

"They are one person
They are two alone
They are three together
They are for each other"
— Stephen Stills

1. Introduction

I'm sitting in my rocking chair on my front porch. One of my sheep, Soren is grazing on the grass on the front lawn. I consciously perceive Soren adequately from this perspective. In consciously perceiving Soren, I am also—though in a thin and minimal way—conscious of myself perceiving Soren. In this instance, as in many familiar instances throughout my day, we might say I am self-conscious. But, how is such self-consciousness possible? There are other creatures that may not be self-conscious in this way, for instance, slugs, squirrels, and maybe Soren. Our evolutionary ancestors may not have been self-conscious in this way. And, presumably, early in infancy, I was not self-conscious. One answer to the question of how self-consciousness is possible is that awareness of one's own mind is *constitutively linked* to awareness of other minds. In other words, being conscious of other subjects is *necessary for* being self-conscious. It is through being aware of other individuals experiencing such and such that I am also aware of myself experiencing such and such. However, such an answer seems dubious if we are gripped by the classic philosophical chestnut—the problem of other minds.

The *philosophical* problem of other minds is raised because of questions like “How do I know that there are any minds like my own?” or “How do I know that other

beings have mental states, such as perceptions, intentions, emotions, thoughts, or experiences in general?” The question is usually raised as an epistemological problem, namely a problem of how we are *justified* in believing that others have minds or how we are *justified* in believing that others have experiences like ours. The assumption that forces us into the problem is that we are surely justified in holding beliefs about our own minds through a *direct, immediate, and transparent way*, but we can only justify beliefs about other minds through *an indirect, mediated, and opaque way*. This has been called ‘the epistemological problem of other minds’ (Avramides, 2001, p. 218f).

But, the questions that generate the epistemological problem assume that our thoughts about our own minds are *direct, immediate, and transparent*. For presumably, to have a belief about one’s own mind is to be disposed to think about one’s mental states in such and such a way. The assumption that we think about our own minds via a direct, immediate, and transparent way assumes that we *always already* possess the concepts MIND, EXPERIENCE, PERCEPTION, ACTION, EMOTION, and other mental concepts. It assumes that all the concepts that enter into our consciousness of our selves and our experiences are given. This gives rise to a problem that has been called ‘the conceptual problem of other minds’ (Avramides, 2001, p. 218f), namely “How is it possible for one to have concepts that apply to minds other than one’s own?” One way to dissolve *the conceptual problem* of other minds is to argue for a thesis that links consciousness of one’s own mind with consciousness of another subject’s mind. (In this paper, I will use ‘consciousness’ and ‘awareness’ interchangeably and will assume that ‘consciousness’ is understood in terms of the disposition to think about the contents of the experience in question.)

In The Paradox of Self-Consciousness, José Luis Bermúdez presents such a thesis and an argument for what he calls ‘*the Symmetry Thesis*’: a subject’s psychological awareness is *constitutively linked* to his awareness of other minds. Bermúdez rightly finds a version of this argument for *the Symmetry Thesis* in the work of Peter Strawson (1959). Both Strawson (1959) and Bermúdez (1998: Ch. 9) understand self-consciousness in terms of the self-ascription of experiences, and I follow them in considering self-consciousness to be the self-ascription of experiences of perceiving, acting and emoting. In Strawson’s now famous essay “Persons,” he offers a version of *the Symmetry Thesis*: in his words, “it is a necessary condition of one’s ascribing states of consciousness, experiences, to oneself, in the way that one does, that one should also ascribe them, or be prepared to ascribe them, to others who are not oneself” (1959: 94) and “I am inclined to think that at least some [transcendental arguments are strictly valid] (e.g., self-ascription implies the capacity for other-ascription)...” (1985: 25). Strawson thought that if we could get clear about the constitutive relation between self and other—in particular, by focusing on the necessary conditions of self-consciousness as self-ascription—then this would enable us to overcome the conceptual problem of other minds.

Bermúdez (1998) provides a needed updating of the Strawsonian argument by bringing Strawson’s transcendental claims into contact with contemporary discussion in developmental psychology. Bermúdez discusses various forms of joint engagement as key phenomena that underwrite his argument for *the Symmetry Thesis*. He argues that the ability to self-ascribe is “constituted” by the intersubjective relations that are realized in joint engagement. (I use the term ‘joint engagement’ for a few reasons. First, in

accounting for the phenomena of joint attention, joint action and joint emotion it is important to accurately characterize the activities between infant and caregiver. Vasudevi Reddy (2005: 104) makes a similar point, suggesting that “attention is conceived here as attending...— as the process through which organisms attend to the world (including their own bodies), rather than as a ‘purely’ mental state that is both discrete and unavailable in action and interaction.” Second, there are conceptual reasons to minimize the attributions to episodes of shared intentionality to capture exactly what infants understand when they are engaging in joint perceiving, joint doing or joint emoting. One of the most difficult tasks in characterizing joint engagements is capturing how the two subjects exploit an understanding of the concept of perception, action and affection (Eilan, 2005: 6). However, it is easier to understand how they both exploit an understanding of engagement. Further, as an anonymous referee pointed out, experience is never solely visual, auditory, olfactory, gustatory, tactile or bodily. Instead, the phenomenology of perceptual experience is multimodal. It is for this reason that ‘joint attention’ also seems too restrictive. (In addition, experience is never solely perceptual, but also includes action and affection. Nevertheless, for the purpose of empirical study, it is necessary to separate out the components of joint engagement.) Joint engagement involves:

- joint attention (infant and adult jointly attending to an object, e.g., a burning candle);
- joint action (infant and adult engaging in an action, e.g., imitation in peek-a-boo); and,

— joint affection (infant and adult entertaining emotions that require intersubjective relations, e.g., pride and shame).

Bermúdez argues that if infants are capable of joint engagement, then infants are capable of other-ascription of psychological predicates. His argument for *the Symmetry Thesis* is the following: infants develop the ability to self-ascribe psychological predicates around 9 months; joint engagement (for example, joint attention, joint action and joint affection) occurs at that time; thus, the best explanation for self-ascription is the intersubjective interaction that is involved in the joint engagement.

I agree with Bermúdez that joint engagement offers an argument and evidence for *the Symmetry Thesis*. However, Bermúdez's account of joint engagement does not provide an argument for *the Symmetry Thesis*. Nor is his account empirically adequate. The fact that Bermúdez's account is a Neo-Lockean account of joint engagement that is committed to the representation of a psychological subject via non-conceptual sortal poses problems for his account.

Bermúdez's account is 'Neo-Lockean' because it is inspired by two central features of Locke's account of the person: (1) Locke's account of the person depends upon a psychological criterion of personhood under which a person is a conscious psychological subject; (2) Locke's account of the person is a reflective account insofar as what determines the person is reflection upon one's experiences. For an interesting discussion of another alternative to the Lockean view of self and person, see the work of John Barresi and Chris Moore (Barresi & Moore, 1996; 2008; Barresi, 1999; 2004; 2010)), much of which is consistent with the view I present in this paper. Barresi and Moore's account is decidedly a persons approach to the self and self-consciousness.

In §2, I argue that Bermúdez’s Neo-Lockean account of joint engagement—since it depends upon a psychological sortal—cannot show that self-ascription is constituted by intersubjective relations in joint engagement (the conceptual critique) and is too rich because grasping sortals—especially a psychological sortal—is a more difficult ability than joint engagement (the empirical critique). However, this paper is not merely negative. Instead of the Neo-Lockean account, I propose in §3, ‘the Persons Theory’ of joint engagement, which proposes that the recognition and acknowledgement of oneself as a person among persons is central to joint engagement. I argue that the Persons Theory can show that self-ascription is *constituted* by intersubjective relations in joint engagement (the conceptual proposal) and is not too rich, because recognition and acknowledgement of persons is not a more difficult ability than joint engagement (the empirical proposal). In §4, I discuss two objections to my argument, and respond to both. Then, in §5, I conclude by considering the import of my argument and account—I suggest that the Persons Theory provides a genuinely new account of mentalizing that differs from *the theory theory* and *the simulation theory*, which I argue is a desirable result given that the debate between those accounts of mental state ascription has come to a standstill. I also consider some further work to be done.

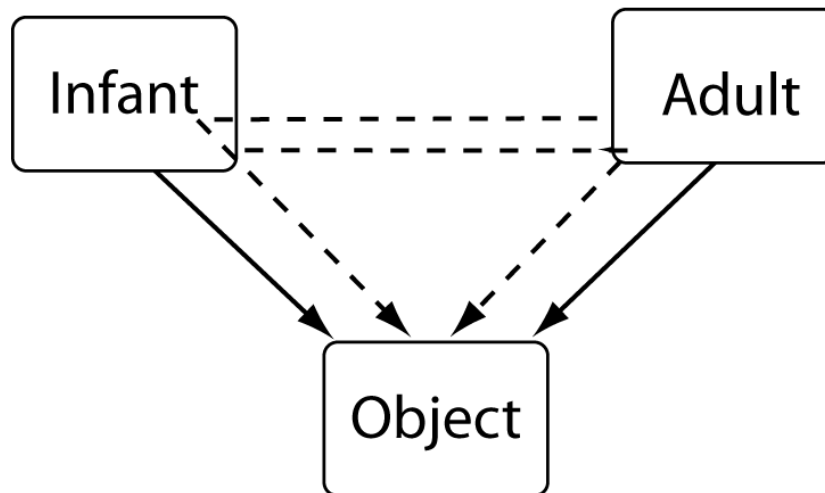
2. The Neo-Lockean Account of Joint Engagement

Before I introduce Bermúdez’s Neo-Lockean account, I will provide some background on joint engagement. At around 9 months, as Piaget (1955) argued, infants begin to undergo a social-cognitive revolution. Infants show the ability to grasp object permanence. Beyond these two developments, infants also begin to engage in joint activities, what Trevarthen and Hubley call “a deliberately sought sharing of experiences

about events and things” (1978: 184). Infants show an increase in the amount of interpersonal engagements. For example, social engagements involved in imitations and other interactions have provided evidence for what might be called “protoconversations” between infant and caregiver.

It is important to note that grasping object permanence and engaging in protoconversations provide evidence for merely DYADIC engagement: (1) interaction between infant and object; (2) interaction between infant and caregiver. However, joint engagement—truly joint, truly shared, truly mutual—is supposed to bring about a transition from DYADIC interaction (2) to TRIADIC interaction, which is a combination of (1) and (2) (Hubble and Trevarthen (1979)) which is usually supposed to take place around nine months. (My focus in the paper is therefore on secondary intersubjectivity—which is characterized by the disclosure of ‘the third element’ in experience, namely, the dawning of a new object of intersubjective engagement—rather than primary intersubjectivity.)

Insert Figure 1: The Triangle: The solid line depicts perception of an object; the dotted line depicts perception of the other perceiving the object. (From Tomasello (1995)).



Caregivers and infants have a distinctive mutual experience. They jointly attend to objects in visual perception. Suppose an infant is on my lap while I am looking at my sheep. In this case, the infant and adult are jointly seeing the sheep. They act in accordance with each other. An infant may point or gaze at a sheep contemporaneously with my pointing or gazing at the sheep. They react similarly to objects in our environment. An infant smiles if the sheep approaches the deck as I smile as Soren approaches. What is relevant for our discussion, in these cases is what is required in order to account for the infant's other-ascribing experiences to me. What does the other-ascription of experiences, for instance, the infant's ascription of my perceiving a sheep, pointing at a sheep, or smiling at the sheep provide in the transition from dyadic interaction to triadic interaction?

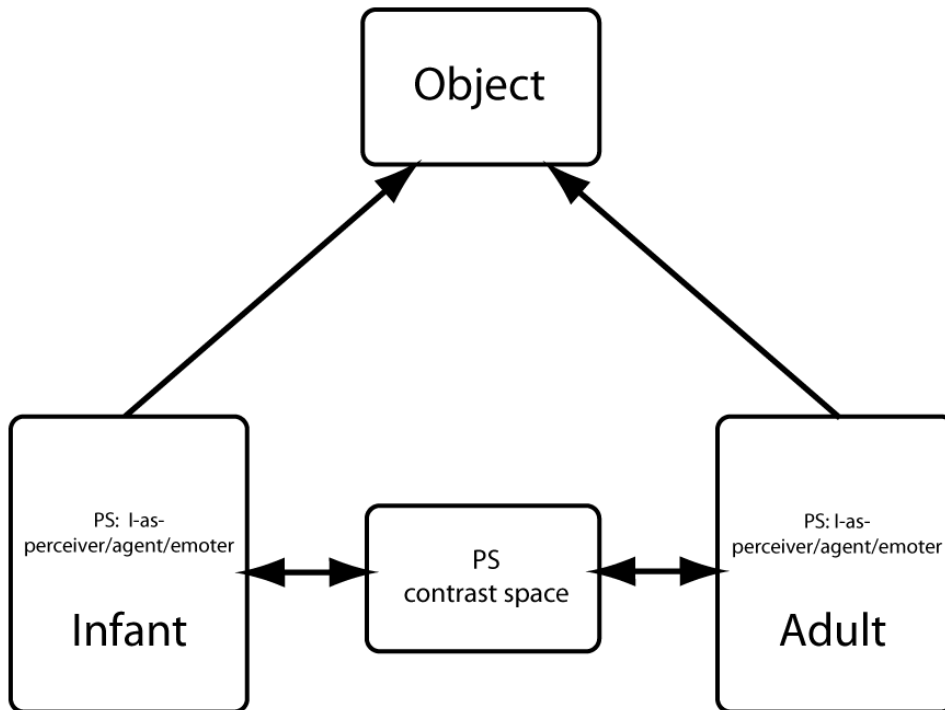
According to Bermúdez, what makes this transition possible is thinking about others and oneself as psychological subjects. Bermúdez presents what he calls ‘a Neo-Lockean account’ of self-ascription in terms of a kind of distinguishing self-awareness. I agree that any account of self-consciousness will involve at least the ability to distinguish between oneself and the world (Cf. Rosenthal, 2005: Ch. 13; Russell, 1995; Bermúdez, 1995, on arguments for the primitiveness of distinguishing awareness for self-consciousness). So, self-ascription of experiences involves such a distinguishing self-awareness. According to Bermúdez, distinguishing self-awareness involves identifying oneself relative to a contrast space: “I have distinguishing self-awareness of myself as Φ only to the extent that I can distinguish myself from other things that are Φ . These other things that are Φ form the contrast space” (1998, p. 238). For example, I might be able to distinguish myself as a creature that sees, and so to form the contrast space, I would distinguish myself from other creatures that see (cf. Russell, 1995, p. 127f). According to Bermúdez, the ways in which a subject can distinguish between self and world requires such a sortal representation of the subject. The requirement of a sortal representation is another reason to consider Bermúdez’s account a ‘Neo-Lockean’ account, because Neo-Lockean accounts of identity involve identification of objects under a sortal (Mackie, 1976, Ch. 5; Wiggins, 1980; Ayers, 1991).

Since self-ascription of “psychological states in the subject-predicate form” (Bermúdez, 1998, p. 233) is “the explicitly psychological dimension of self-awareness” (Bermúdez, 1998, p. 229), then, according to Bermúdez, self-ascription requires discriminating and counting oneself as a psychological subject distinct from the environment. Given that the requirement on distinguishing self-awareness is identifying

oneself relative to a contrast space with respect to a psychological sortal, then the psychological sortal cannot have only a first-person application, but must have third-person application. According to Bermúdez, given the logical structure of the contrast space, and the claim that self-ascription involves such a psychological sortal, then *the Symmetry Thesis* follows—“A subject’s psychological self-awareness is constitutively linked to his awareness of other minds” (Bermúdez, 1998, p. 230).

What determines that the psychological sortal applies to one individual rather than another? In other words, what enters into the content of the representation of a psychological subject? According to Bermúdez, the psychological sortal is a complex sortal, analyzable in terms of more basic categories. The psychological sortal is not a “set of necessary and sufficient conditions that will settle in any case whether one is dealing with a psychological subject or not” (Bermúdez, 1998, p. 243) and “All of the concepts involved here are complex and vague in ways that make it extremely unlikely that they will each always have a determinate criteria of identity” (Bermúdez, 1998, p. 247). However, according to the Neo-Lockean account, the sortal is determined by strong *prima facie* evidence: “a judgement based on such evidence would count as warranted..., a judgement that would often settle a debate as to the presence of [a psychological subject] and would at the minimum be taken very seriously in such a debate” (Bermúdez, 1998, p. 243). So, according to this account, the psychological sortal is complex, vague, and dependent upon philosophical debate. What enters into the complex sortal? Bermúdez settles on the following principle: “(PS₂) Psychological subjects with a perspective on the world are aware of themselves as perceivers, agents, and as having reactive psychological states” (Bermúdez, 1998, p. 247). See figure 2.

Insert Figure 2: Bermúdez's Neo-Lockean Account of Joint Engagement.



Two critical questions arise about the Neo-Lockean account of joint engagement. The first question is, “Can this account of joint engagement provide an argument for a constitutive link between self-ascription and other-ascription?” Remember that the argument for *the Symmetry Thesis* depends upon this. The second question is, “Are the representations that Bermúdez attributes to the infant proper to its cognitive abilities at 9 months?” Since joint engagement occurs around 9 months, any representations attributed to the infant should be proper to its abilities.

I begin with a couple caveats. My critique of the Neo-Lockean account is not meant to suggest that joint engagement in general cannot provide an argument for *the*

Symmetry Thesis since, as I argue below, that it can. Instead, I take issue with the description of joint engagement in terms of a psychological sortal. Also, it is well known that Bermúdez’s rich, unique and interesting account of self-consciousness provides arguments for non-conceptual forms of consciousness of self. Although, it might seem unintuitive that a sortal could be non-conceptual, given that a sortal is generally understood to be a representation that enables classification under a category and categorization is also generally understood to be central to concept possession, my critique will not depend upon the question of whether the psychological sortal is conceptual or non-conceptual, since I take the debates about non-conceptual contents to be orthogonal to the issue at hand.

To answer the first question, we have to demystify the notion of a constitutive link. What does it mean to provide an argument for a *constitutive link* between self-ascription and other-ascription? One historical argument for a constitutive link is Kant’s (1781) argument for the link between experience of an object and the categorization of objects in space and time. The inspiration for Kant’s argument arises because of the question, “What make possible experience of an object?” Kant’s answer provides an argument for what is necessary for the experience of an object, namely the categories of thought and experience and the pure intuitions of space and time. I am not concerned with Kant’s argument itself here, but instead with the structure of the argument. The question about a *constitutive link* between self-ascription and other-ascription is “What makes possible the self-ascription of experiences?” I am suggesting, on analogy with Kant’s argument, that we should provide an argument that what is necessary for the self-

ascription of experiences is the other-ascription of experiences. This alone will not clarify how this constitutive link might relate to developmental psychology.

However, a quote from Johannes Roessler provides a further empirical articulation of the notion of a constitutive link: "...the task facing developmental psychologists is not that of mapping the kinds of evidence on the basis of which infants attribute mental states to others, but that of giving an account of the way in which infants' perceptual experience becomes imbued with psychological concepts. It is here that social interaction may have a constitutive role to play" (2005: 252; my underlined emphasis). The claim is that the concepts that enter into self-ascription are determined by the acquisition of those concepts in social interaction. I will return to Roessler's account of joint engagement below. For now, the aim is to make sense of how joint engagement provides for a logical (transcendental) dependence between self-ascription and other-ascription, such that the same representations that enter into one's awareness of oneself and one's experiences can (*in principle* at least) also enter into one's awareness of an other and an other's experiences.

I will argue that the Neo-Lockean account cannot meet this aim, because it cannot be used to argue for a constitutive link. The constitutive claim cannot go through because, even according to Bermúdez's interpretation of 'constitutive' as logical dependence (Bermúdez, 1998, p. 231), the application of a psychological sortal to oneself must be logically dependent upon the application of the same psychological to the other, and vice versa. However, as we saw above, the psychological sortal is defined by the principle "(PS₂) Psychological subjects with a perspective on the world are aware of themselves as perceivers, agents, and as having reactive psychological states"

(Bermúdez, 1998, p. 247; my underlined emphasis). Since, by Bermúdez’s definition, this is the sortal that enters into self-ascription, the criterion that is used for the psychological sortal is (*in principle*, at least) something that can apply in its first-person applications in a way that diverges from its third-person applications. So, the contrastive dimension that Bermúdez alludes to implies a contrast between the first-person and the third-person that leaves an unbridgeable gap between the self and other. Insofar as it leaves this gap in place, therefore, it cannot be used to argue for a constitutive claim, and so it cannot be used to argument for *the Symmetry Thesis*.

To answer the second question—“Are the representations that Bermúdez attributes to the infant proper to its cognitive abilities at 9 months?”—we need to look closer at the psychological sortal, putting aside its logical properties, and instead consider the question of whether an infant at 9 months could feasibly entertain such a representation. However, the kind of psychological sortal that Bermúdez requires is far too robust. The grasping of sortals is more complex than self-consciousness itself (Xu, 2007). It is not until later than 12 months that infants begin representing such simple objects as ducks and balls with sortals, never mind perceivers (Baker, 2003; Xu, 2007). Bermúdez argues however, that representing a contrast space between oneself and the other via a sortal is required in order for self-ascription to occur, as the diagram above suggests. In that sense, he thinks that that is what is required for joint engagement as well. What this means is that in order for an infant to enter into joint engagement, the infant needs to represent the caregiver as a perceiver and represent itself as a perceiver, where it is assumed that a perceptual content, for instance a sheep, is the object of joint engagement. However, this clearly requires an iteration of representations, because in

order for joint engagement to count as *joint*, it also needs to represent these two facts, which also requires representing itself as a perceiver, as I discussed above.

These two critiques show just how difficult it is to account for joint, shared, mutual or genuinely intersubjective experiences in a way that ties self and other together such that it provides an argument for the constitutive link. Further, it also makes it difficult to theorize about joint engagement in a way that does not require too much of infants, for instance by attributing to them attitudes that only adults could possess. Allow me to offer a diagnosis. The problematic tendency is to think that joint, shared, or mutual experiences are a complex of two fully-formed subjectivities, tied together by some further, more mysterious and objective relations. In the case of the Neo-Lockean account, the commitment to a psychological sortal involves the commitment to attributing to the infant such a fully-formed subjectivity. And, the notion of the contrast space is supposed to capture the objective meeting-of-the-minds between infant and adult. However, as I will argue, there is little reason to be committed to these assumptions. If the claim about a constitutive link is correct, then the Hegelian claim that “the I is a We; the We, an I” should not commit the infant to possessing a fully-formed subjectivity, but instead the intersubjective relation between infant and adult should be understood as a joint, shared and mutual subjectivity. I will argue that a more minimal account of what the infant represents provides for an argument for the constitutive claim and better captures the phenomena. Further, my account can be considered to be legitimately objective as well, objective in the sense that “these activities are conceived as being independent entities in a world that is independent of the representer” (Grush, 2002). I will take the work of Peter Strawson as an inspiration in developing this account, as his

discussion in “Persons” provides pivotal insights into joint engagement and the development of self-consciousness.

3. The Persons Theory of Joint Engagement

In The Bounds of Sense, Strawson discusses the idea that self-ascription and other-ascription depend upon a ““determinate intuition” corresponding to the concept of a subject of experience”... which requires “that there be empirically applicable criteria of identity for subjects of experience. In actual practice this condition is satisfied by the fact that each of us is a corporeal object among corporeal objects, is indeed a man among men” (Strawson, 1966, p. 102; my underlined emphasis). Strawson also argues that in order to account for self-ascription and other-ascription, we have to presuppose “the primitiveness of the concept of the person” (Strawson, 1959, p. 97). It is my contention that an account of joint engagement can be developed that takes the notion of the person as central. One psychologist concerned with joint engagement, Peter Hobson (2002), has argued that what is required for joint attention is “intersubjective engagement” (Hobson, 2002, p. 201) *between persons*. I discuss Hobson’s account below in more detail. (Another recent discussion of the importance of persons is Newen and Schlicht’s (2009) Person Model theory.)

Two clarifications are required in articulating the Persons Theory of joint engagement. Strawson does not intend to shift the debate about self-ascription and other-ascription to a debate about *personal identity*, and neither do I. That would over-intellectualize the concept of the person as it functions in joint engagement. Following an underrepresented tradition in philosophy and psychology, I take the concept of the person to be a primitive. (For a helpful discussion of the primitiveness of the concept of

a person with respect to the relation between self and other, see John Haugeland (1982), Peter Hacker (2001), Marjorie Grene (2002) and John Barresi (1999).). Also, although Strawson himself agrees with Ishiguro's (1980) interpretation that "in order to *be* a person, one must see oneself and others *as* persons. One must have an operative concept of a person..." (Strawson, 1994, p. 214), a Persons Theory of joint engagement need not be committed to attributing concepts or representations to the infant, since as I will argue recognition and acknowledgement of persons is leaner than that.

Strawson's explicit concern in "Persons" accounting for how experiences are ascribed to one and the same individual that possesses both mental and corporeal characteristics, and which can be *identified* independently of those characteristics. While there are many ways by which a subject may identify itself, Strawson suggests that in order to account for self-ascription and other-ascription, we should privilege the concept of the person. But, there is a clear difference between Strawson's view of experience ascription and Bermúdez's view of experience ascription. I will provide an articulation of John Campbell's (2002) helpful distinctions concerning joint engagement in order to make an initial distinction.

According to Campbell's analysis of joint engagement, there are two ways to distinguish between different accounts of joint engagement. The first distinction is between an experientialist account and a non-experientialist account. According to an experientialist account, there is an experiential difference between perceiving a sheep and jointly perceiving a sheep. I might sit on my porch perceiving a sheep. My wife sits down next to me and says, "Soren needs a shave" so that I move from solitary perceiving of a sheep to coincidentally co-perceiving a sheep, to full joint perceiving of Soren.

According to the non-experientialist account, there is no experiential difference between perceiving a sheep and jointly perceiving a sheep. According to the non-experientialist account, the jointness, sharedness or mutuality is *external* to the experience of the individual. Both Bermúdez's and Strawson's accounts are experientialist accounts.

However, Campbell makes a distinction within experientialist accounts. There are those experientialist accounts that provide relational analyses and those that provide reductive analyses. A reductive analysis of joint engagement posits the mental states of an individual that are relevant to joint engagement without already implying that there is joint engagement involving one individual and another individual. However, on a relational analysis of joint engagement, this is not possible, because to posit the joint engagement of one individual is to *ipso facto* posit the other as a constituent. As John Campbell suggests that on a relational view, joint engagement is a primitive phenomenon of consciousness (2002: 288). According to these distinctions, Bermúdez's Neo-Lockean account is a reductive account of joint engagement, because the psychological sortal is a representation entertained by each individual independent of the contrast space between them. The contrast space is composed of the distinct representations in distinct spheres—individual minds. However, the Persons Theory of joint engagement is an experiential relational view of joint engagement that does not require positing two distinct self-standing subjectivities.

Recently, taking up Strawson's claim and also offering an experiential relational view, Axel Seemann (2008) argues that a Strawsonian account of joint engagement requires providing an account of what he calls 'person perception.' I will discuss

Seemann's account in light of my conceptual critique of Bermúdez above, since his view and mine are similar.

In order to defend his account of person perception, Seemann appeals to Hobson's account of feelings between infants and caregivers. The view involves the infant and the caregiver related through "feeling perception": "there is a mode of feeling perception that is critical for establishing intersubjective relations between people, and it is a kind of perception that establishes a special quality of relatedness between the individual and what is perceived— in most natural circumstances, a person" (Hobson, 2005, p. 199). Feeling perception involves the infant and the caregiver engaging in perspective taking, where perspectives are shared. According to Seemann, "the intersubjective perspective enables an infant to develop an understanding of the distinction between self and other that then makes it possible for him to engage in this kind of perspective-taking. It is precisely the *sharing* of feelings that puts one in a position to think of oneself and other persons *as selves*" (Seemann, 2008, p. 250). So, according to what might be called 'the feeling perception account,' what makes possible joint engagement is not as robust as Bermúdez's account via a psychological sortal, because there is no reason that feeling perception need be committed to concepts or representations at all, as Seemann suggests. However, there is a concern about whether this account can meet the Strawsonian requirements that Seemann sets out.

According to Seemann, any account of person perception needs to capture how we engage in jointly experiencing (meeting "the Symmetry requirement") while maintaining a separation between bodies (meeting "the Embodiment requirement") (Seemann, 2008, pp. 246–7). I am most interested in the Symmetry requirement, which

suggests that the attribution of psychological states to the other cannot be modeled on the attribution of psychological states to oneself alone. For example, if an infant is to attribute to the caregiver that he fears a sheep, then that attribution cannot be based solely on the infant's self-ascription of fear in the face of the sheep. If that were the case, then first-person ascription of psychological states would be prior to second- or third-person ascription of psychological states.

In an attempt to avoid an account that supposes a first-person priority, Seemann develops an account of joint engagement that is based on the shared experiential meaning of that he calls "simple feelings." According to this account, an infant acquires an ability to affect the caregiver through multiple experiences of the effects of expressing a simple feeling. Suppose that the infant expresses happiness with a smile to her caregiver. The social environment is such that it makes possible the perception of similar simple feelings in the other. "When I smile at another person, I do not only express an aspect of my subjective life: I also affect that person's psychology, and I am immediately aware of that fact" (Seemann, 2008, p. 258). This provides a type of embodied know-how of the causal role of such simple feelings. In tandem with this development, one's own simple feelings are affected by the other's simple feelings, which provides an understanding of oneself as a part of a social environment. Through the reciprocal effects of one's expression of simple feelings and the other's expression of simple feelings, the infant grasps that one is "causally on a par with other members of that environment" (Seemann, 2008, p. 258).

Seemann suggests that this account satisfies the Symmetry requirement, because "it maintains that self- and other-understanding develop conjointly" (Seemann, 2008, p.

258). While I agree with Seemann's account overall, it seems the feeling perception account is not sufficient to meet the Symmetry requirement as expressed by Strawson. According to the feeling perception account, the infant has the capacity to self-ascribe simple feelings before it has the capacity to other-ascribe simple feelings. In this sense, then there is an asymmetry between self and other in this account. While the feeling perception account does not commit to the psychological sortal that the Neo-Lockean account does, it nevertheless commits an infant to the self-ascription of simple feelings independent of the other-ascription of simple feelings. Thus, first-person and second-person ascription can come apart.

Against both the Neo-Lockean view and the feeling perception account critiqued above, I propose an account of person perception in terms of the recognition and acknowledgement of persons, which I call the Persons Theory. First, this account can show that self-ascription is constituted by intersubjective relations in joint engagement, meeting the conceptual critique. Second, the Persons Theory is not too rich, because recognition and acknowledgement of persons is proper to infants' cognitive abilities at 9 months. Third, the Persons Theory does not focus on the intersubjective intentionality of simple feelings as the means of connecting the self and other, but instead, focuses on embodied interaction between individuals.

I will elucidate the Persons Theory by first outlining a Strawsonian argument for the Symmetry Thesis, since the constitutive claim is the conclusion of that argument. According to Strawson, the ability to self-ascribe requires the ability to other-ascribe, and vice versa. This transcendental claim can be rephrased in empirical terms as the

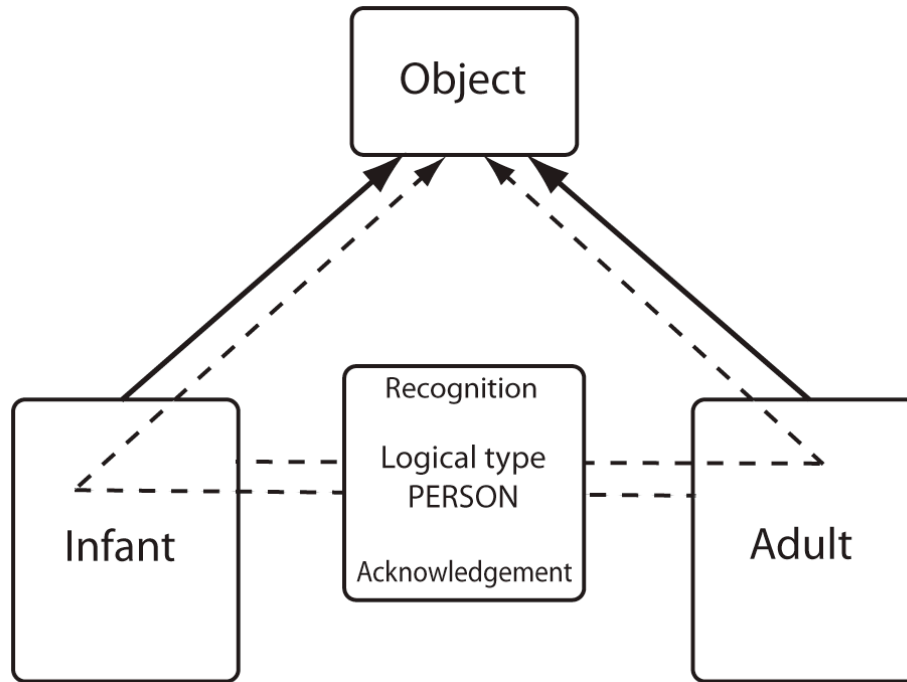
hypothesis that self-ascription and other-ascription develop conjointly or in tandem. The argument from “Persons” is as follows:

1. Self-consciousness should be understood in terms of the self-ascription of experiences, as proposed by many philosophers and psychologists (Kant, 1781; P. F. Strawson, 1959, 1966; Evans, 1982; Brook, 1994; Cassam, 1997; Bermúdez, 1998).
2. The self-ascription of experiences requires meeting the generality constraint: if a subject is to be able to self-ascribe the experience “I am F,” then she must have two distinct abilities: (GC₁) she must be able to ascribe experiential predicates, e.g., “b is F,” “c is F,” to arbitrarily distinguishable individuals; (GC₂) she must be able to ascribe “I am G,” “I am H,” for any experiential predicates she possesses independent of “I am F.”
3. In order for self-ascription to meet the generality constraint, the subject must be able to identify itself as a subject of experience; in other words, self-ascription requires self-identification.
4. Self-identification is achieved through the perception of oneself and others as persons among persons.

Thus, according to Strawson, self-consciousness is made possible by the ability to perceive oneself as a person among persons. I will not focus on premises 1-3, but instead will assume them here. What is involved in recognition and acknowledgement that enables us to ascribe experiences to ourselves and to others? Notice, this account of self-consciousness and other-consciousness depends—as the simple feelings account does—on an account of person perception. However, the Persons Theory differs from the

simple feelings account because it focuses on recognition and acknowledgement as central to person perception, rather than the perception of simple feelings. For example, when I think, “I am seeing a sheep” and when I think “You are seeing a sheep” (or “Taylor is seeing a sheep”), then the first-person, second-person (or third-person content) of those ascriptions is determined by ascription under a logical type of subject, namely the logical type <person>. (By ‘logical type,’ I mean the commonsense category under which individuals fall given the set of activities in which those individuals engage. Strawson’s (1966) account of the person focuses on the logical type person instead of focusing on the dimensions of ‘the mind’ or ‘the body’, because the role of the entire organism in self-consciousness is prior to the roles of the dimensions of mind and body.) Instead of a psychological sortal that each subject represents oneself and the other with, the activity of interpersonal engagement itself embodies the logical type <person> (Figure 3).

Insert Figure 3: The Persons Theory of Joint Engagement



Above, I said that the Persons Theory focuses on the claim that self-identification is achieved through the perception of oneself and others as persons among persons. On my account, perception is understood in terms of recognition and acknowledgment. Because ‘recognition’ and ‘acknowledgement’ are terms that are original to my account, I will explicate them. (Also, while the Persons Theory employs the terms ‘recognition’ and ‘acknowledgement,’ the Persons Theory should not be assimilated to Axel Honneth’s Hegelian account of self-other relation (1996).) Recognition involves perceptual-based identification of macroscopic activities and properties of individuals. Acknowledgement involves manifesting an understanding of what Roessler has called “interpersonal normative relations” (Roessler, 2002, p. 249). The process of recognition and acknowledgement goes through three stages: sharing engagement, following engagement, and directing engagement (cf. Carpenter et al (1998)). I will elucidate recognition,

acknowledgement and the stages of engagement. The Persons Theory of joint engagement depends upon recognition and acknowledgement between the self and other. My claim is that recognition—the infant’s classification of the adult as like itself with respect to its activities and embodiment (recognition should be understood in terms of the participatory features of joint engagement)—and acknowledgement—the adult’s initiation of the infant into the activities and practices that are typically human (acknowledgement should be understood in terms of the negotiatory features of joint engagement)—are constitutive of the contents of experiences ascribed to self and other. Are there more particular claims that can be made about recognition and acknowledgement so that the account may be tested in light empirical evidence? Although my intention in this paper is mostly to lay out the Persons Theory, as I make recognition and acknowledgement more clear and precise, I provide empirical predictions of the Persons Theory.

The Persons Theory of self-ascription and other-ascription would, thus, need to account for how we directly perceive persons and account for the various person-predicates—perceiving, doing, emoting—that enter into self-ascriptions. In “Persons,” Strawson suggests that insofar as we recognize ourselves and others as perceiving, then we must have the ability to recognize creatures with a sensibility, a location and an orientation (Strawson, 1959, p. 99). This passage is often ignored in discussion of “Persons,” but it is essential to a Strawsonian account in general, and will be pivotal to the Persons Theory. It also provides us with hypotheses that may be developed further to provide experiments that could test the Persons Theory. For instance, results in the gaze-following paradigm (Tomasello *et al*, 1998) suggest that primates follow the gaze

direction of conspecifics and human beings dependent upon recognition of like sensibility, geometrical location and orientation in space. One test would be to determine whether primates are capable of being aware of themselves as perceivers when they are not capable of recognition of the other's possessing a sensibility, geometrical location or orientation. Does recognition of the fact that other primates have eyes, ears, noses, or other macroscopic properties of primate sensibility contribute to their own ability to be aware of themselves as perceivers? It has been argued that primates' gaze-following is in many ways similar to human infants' gaze-following (Butterworth and Jarrett, 1991). Insofar as the focus is on the self-ascription of perception, the Persons Theory would predict that those infants that are not capable of recognition of other person's sensibility, geometrical location or orientation would have deficits in the development of the self-ascription of perceptual experiences. Within the gaze-following paradigm, Brooks & Meltzoff (2002) provide evidence that the infant's recognition that the adult's eyes are open (rather than merely covered by a blindfold) is significant for gaze-following. More experiments could be done on infants younger than 12 months that focus also on modalities other than vision as well. The Persons Theory would predict that recognition of features of sensibility, location and orientation are relevant for the self-ascription of perceptual experiences.

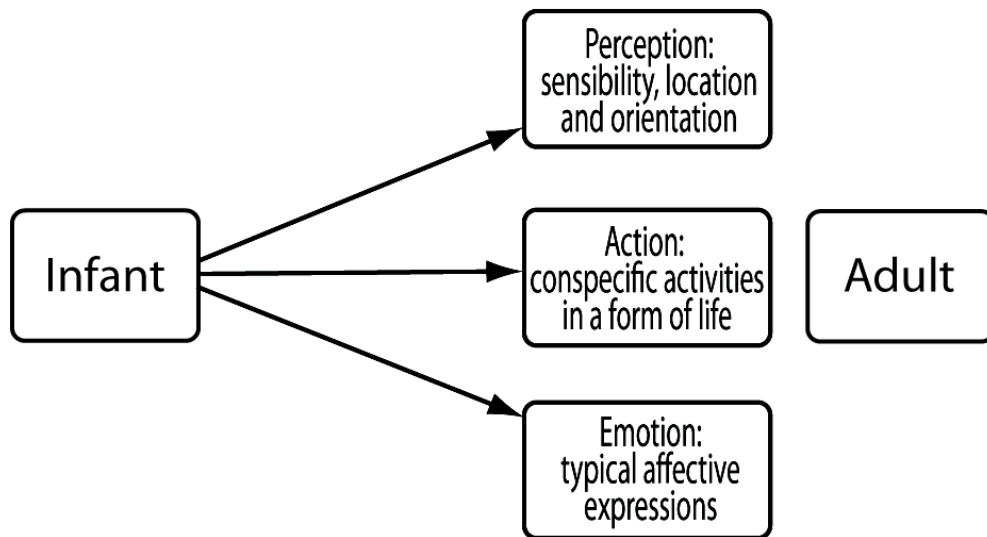
About the recognition of action, Strawson suggests: "it is easier to understand how we can see each other, as persons, if we think first of the fact that we act, and act on each other, and act in accordance with a common human nature" (1959, p. 109). The links between the infant's body and the caregiver's body enables forms of interpersonal recognition of movements. It has been argued that there are supramodel body schemata

that enable infants to recognize symmetries between herself and other persons (Meltzoff and Moore, 1994; Meltzoff, 2007; Gallagher and Meltzoff, 1996): “the body schema, working systematically with proprioceptive awareness, operates as a proprioceptive self that is always already ‘coupled’ with the other” (Gallagher and Meltzoff, 1996, pp. 225–226). Insofar as the focus is on the self-ascription of action, the Persons Theory would predict that if infants lacked recognition of the adult’s bodily movements as actions of persons, then infants would have deficits in the ascription of actions to themselves. For example, consider an infant that could not distinguish between intentional bodily movements of persons (with a target and purpose) and unintentional bodily movements of persons (without a target and purpose). Does such an infant display a parallel lack in the capacity self-ascribe its own bodily movements as actions? The Persons Theory expects that such deficits would occur.

About the recognition of emotions, Strawson suggests: “X’s depression *is* something, one and the same thing, which is felt, but not observed, by X, and observed, but not felt by others than X” (1959, p. 108). It is clear from these quotes that according to the Persons Theory, recognition involves the direct perception of the contents of experiences—perceptions, actions and emotions—asccribed to self and other. Insofar as the focus is on the self-ascription of emotions, the Persons Theory would predict that if infants lacked recognition of adult’s affectation, then infants would have deficits in the ascription of emotions to themselves. Researchers might test whether there are correlations between a failure to ascribe emotional states to individuals within the point-light experiments (Johansson, 1973; Bertenthal, Proffitt, and Cutting, 1984; Hobson, 2005) and failures of ascribing emotional states to oneself.

I represent recognition with Figure 4.

Figure 4: Recognition: Lines represent perceptual recognition of activities on the part of the adult.



The Persons Theory of joint engagement can also account for the normative aspect of joint engagement by providing an account of acknowledgement. In the discussion of ascription, Strawson argues that the predicates that enter into self-ascriptions and other-ascriptions do not “have two kinds of meaning. Rather, it is essential to the single kind of meaning that they have, that both ways of ascribing them should be perfectly in order... the predicate gets its structure from the whole structure of the game” (1959, p. 107). While this game metaphor may seem to be merely a Wittgensteinian flourish of the pen, it offers a way of capturing a pivotal aspect of joint engagement that has been a question mark recently. What is essential for the notion of

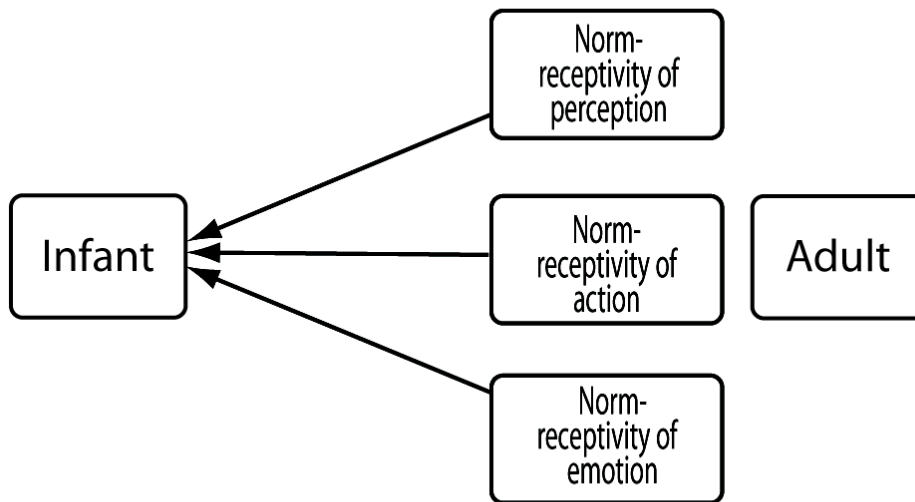
acknowledgment is the normative character of the relation between infant and caregiver involved in joint engagement. Joint engagement is *not merely* a causal recognitional ability of one's conspecifics, but also involves an interplay between self and other. Johannes Roessler has argued that what is significant about joint engagement is that "interpersonal causal relations assume a normative significance" (2002, p. 249). According to Philip Rochat, "it appears that from 9 months of age, infants become explicitly aware of others as potential teachers and informants... In particular, infants construe others as having authority with the power to judge because they know better" (Rochat, 2001, p. 356).

As Roessler suggests, it might seem too rich to suppose that an infant can grasp the normative significance of the embodied engagements with the other that are constitutive of the content of experiences such as perceptions, actions and emotions. After all, infants do not yet possess the capacity to give and ask for reasons, namely they do not express entitlement or commitment to justifications. Further, they cannot express in language that the caregiver is the authority concerning understanding of perception, action and emotion. But, there is a much weaker sense of normativity that enters into acknowledgement.

Much of the discussion about joint engagement—joint attention, joint action, joint affection—focuses on the transition from dyadic interaction to triadic interaction, as I discussed above. An observation that is central to this transition is the role of interactive play between infant and caregiver (Kagan, 1998). Often, we think of play as something that is without rules and regulations. This, however, is not the way to think of the type of play between infant and caregiver. The transition from dyadic engagement to triadic

engagement involves a transition from (1) the acknowledgement of the other's authority where perceiving, acting and emoting is concerned to (2) taking up a role or function in the game of perceiving, acting and emoting, and thereby learning the correctness and incorrectness conditions of those activities. Thinking of joint engagement as a structured game captures the normative significance of acknowledgement. I represent acknowledgement with Figure 5.

Figure 5: Acknowledgement: Lines represent normative acknowledgement of activities on the part of the infant.



According to the Persons Theory's account of acknowledgement, if there are infants that are not capable of engaging in interactive play, then we should find parallel deficits in the same infants' capacity for the self-ascription of experiences. While going into the

details would take us too far afield, Peter Hobson's research (summarized in 2002) with autistic infants and children might be read as suggesting that interactive play is essential to the dimension of acknowledgement. Autistics like 'L' seem to have a lack of engagement in interactive play with caregivers and show parallel deficits in self-ascription.

To return to the issues raised for Bermúdez's account above, we might ask, "How does the Persons Theory support the Symmetry Thesis?" According to the Persons Theory, when I think, "I am seeing a sheep" and when I think "You are seeing a sheep" (or "George is seeing a sheep"), then the first-person, second-person (or third-person) content of those ascriptions is determined by ascription under a logical type of subject, namely, the logical type <person>. According to the Persons Theory, the criterion for the identification of the person is public and is taken up in the practice of recognizing and acknowledging each other as persons under that logical type. So, there is no possibility of a logical independence between self-ascription and other-ascription. Since there is no logical independence between self and other, then the constitutive claim can go through and we can make sense of the claim that self-ascription and other-ascription are reciprocally dependent. But, what about the issue of whether the Persons Theory places too much demand upon an infant at 9 months? (I am grateful to an anonymous referee for pushing this point.)

Does the recognition and acknowledgement of persons commit us to a rich account of joint engagement? In order to account for the recognition and acknowledgement of persons, all that is necessary is positing practical know-how, such as embodied skills. The recognition and acknowledgement of persons does not require

grasping the sortal of a psychological subject. However, there is evidence that infants can grasp the sortal <person> at around 9 months (Bonatti *et al* (2002); Xu (2007)).) One might object that recognition and acknowledgement seems rich, because it is not only that an infant perceives the adult's sensibility, location, orientation and that the adult acknowledges the infant through passing on the norms of perception, action and emotion, but that such embodied activities are linked to consequences like perceiving, acting and emoting as such. According to this objection, the infant has to entertain *inferences* from embodied activities to thought about such embodied activities. However, the Persons Theory (following Strawson's lead) resists such an intellectualization of intersubjectivity.

Peter Hobson (2002) argues that what is required for joint attention is "intersubjective engagement" (201): "One can only have joint attention if one has the capacity to 'join' another person—which means that one needs to be able to share experiences with others, registering intersubjective linkage—and at the same time remain separate" (201). In order to achieve separation, such intersubjective linkage need not be understood as based in *thoughts about* or *imaginings of* others; instead, as I argued above, perception of others is sufficient to account for the intersubjective linkage without over-intellectualizing intersubjectivity. The Persons Theory accomplishes both of Hobson's requirements, because it captures how we engage in jointly perceiving (meeting "the Symmetry requirement") while maintaining a separation between bodies (meeting "the Embodiment requirement") (cf. Seemann (2008: 246–7)).

How might the Persons Theory joint engagement provide for an account of ascription of experiences? According to the Persons Theory, both the caregiver and the infant draw upon perception of persons to enable them to ascribe experiences. So, the

infant recognizes and acknowledges that a person is perceiving not because it *represents* an other as a perceiver or because it *represents* itself as a perceiver through a psychological sortal. Instead, it *directly* perceives that other persons are engaging in activities that manifest perception, action and emotion. As I discussed above, there are a variety of manifestations of persons perceiving that are easily grasped. For instance, a subject's gaze is a manifestation of seeing, a subject's turning her head is a manifestation of hearing, a subject's bodily movement is a manifestation of an intentional action and a subject's grimace is a manifestation of anger. These are all typically personal level activities that the infant can directly perceive and grasp as manifesting perceiving, acting and emoting. It is important to note here that Strawson was not inclined to accept that activities such as these were merely behaviors, since it is not solely mental attributions that determine that perceptions, actions and emotions are intentional. Instead, for him they are directly perceived as intentional personal level activities.

4. Objections to the Persons Theory

I will consider two objections to the Persons Theory. The first objection contends that there is no real difference between the Neo-Lockean account and the Persons Theory. The second objection charges that I have presented little evidence for the relevance of the logical type <person> for ascription of experiences to the individual infant, which is necessary in order to argue for *the Symmetry Thesis*. (I am grateful to John Barresi for proposing this objection.)

It might be argued that there is no real difference between the Neo-Lockean account and the Persons Theory, because both accounts involve the attribution of the person predicates of perception, action and emotion. Although this is true, there are

significant differences between the two views. First, as I discussed above, the Neo-Lockean theory is a reductive theory and the Persons Theory is a relational theory. Second, the Persons Theory does not commit to the *representation* of self and other, because recognition and acknowledgement is theorized as an activity for which embodied skills and abilities (mostly perceptual skills and abilities) are sufficient. Since there is no need to appeal to representations in the Persons Theory, there is no threat of an iteration of representations as there is with the Neo-Lockean account. Third, the Persons Theory is different because it puts aside the issue of whether there are psychological versus non-psychological predicates, and correlatively, whether the generality constraint requires ascription of experiences to individuals distinguishable as psychological subjects. This is important to register, given that much of the literature on self-ascription makes the distinction between mental self-ascriptions and bodily self-ascriptions. However, Strawson's concern in "Persons" is to account for how experiences are ascribed to one and the same individual that possesses both mental and corporeal characteristics. On this account, generalizing the subject of self-ascription depends crucially on identification of the subject of experiential predicates independently of the distinction between the mental dimension and the bodily dimension (For a similar account, see Barresi, forthcoming.).

The second objection suggests that I have not adequately defended the claim that the Persons Theory can be used to meet the Symmetry requirement. While I provided evidence through the Bonatti (2002) study that infants can operate with the sortal person in ascription to other subjects, I have provided little evidence that the infant is able to apply that sortal to itself. But, in order to provide for symmetrical relations between self

and other, I must provide evidence that an infant can apply the logical type <person> to itself.

My reply is that there may be evidence from within the mirror self-recognition paradigm that infants are capable of ascribing the sortal <person> to themselves (Gallup (1970)) insofar as they are able to pass a version of the mirror test. I grant that children treat the other baby in the mirror as a different individual until around 2 years old, but there may be new measures of mirror self-recognition developed that enable us to find recognition of person properties earlier—which would provide a framework for a person-recognition test. A more recent study that may be used as a springboard for future research is Lewis and Carmody's (2008) discussion of the correlations between the capacity for mirror self-recognition and maturation in brain development. While Lewis and Carmody's studies focus on young children 15-30 months, the Persons Theory would predict that both infants around 9 months would be able to pass a person-recognition test and that there would be differences in experienced-based maturation in brain development before and after the capacity for self-recognition as a person. In addition, some progress might be made in considering results from deferred imitation studies because deferred imitation develops around 9 months (Meltzoff, 1988). In cases of deferred imitation, the infant needs to remember the movements of the adult in such a way that is best explained in terms of representing its own action under the sortal <person>. Addressing this objection completely requires more development of the Persons Theory and more empirical research.

5. Concluding Remarks

According to the Persons Theory, joint engagement involves two individuals perceiving an object through sharing perceptions, actions and emotions. If Strawson is right that self-ascription and other-ascription is done under the logical type <person>, then Hegel might have been right to say “the I is a We; the We, an I.” On this view, the ‘we’ is not composed of two self-standing subjectivities; joint engagement is not two minds alone meeting somewhere in the middle. What is the import of this for accounts of mentalizing or mindreading? I can only touch on this issue briefly here. The two classic accounts of mindreading are the theory-theory and the simulation theory. The theory-theory, in a nutshell, holds that the ascription of experiences is made possible by the possession of a theory of other minds. The simulation theory claims that the ascription of experiences is made possible by the simulation of other minds. The Persons Theory asserts that *the ascription of experiences is made possible by the perception of other minds*.

The Persons Theory is a genuinely new theory. Much of the discussion about mindreading is whether the theory-theory and the simulation theory collapse into each other (Davies (1994) and Heal (1994)). The theory-theory (TT) and the simulation theory (ST) both share the fact that they are asymmetrical views, whereas *the Persons Theory* is a symmetrical view. There are three possible views of mentalizing:

(1): *self-priority asymmetry*: the ability to self-ascribe is not dependent upon the ability to other-ascribe in the sense that one need not think about or perceive others in order to self-ascribe (ST);

(2): *other-priority asymmetry*: the ability to other-ascribe is not dependent upon the ability to self-ascribe in the sense that the ability to other-ascribe is prior to the ability to self-ascribe (TT);

(3): *no-priority symmetry*: the ability to self-ascribe is mutually reciprocally dependent upon the ability to other-ascribe in the sense that one cannot self-ascribe unless one can other-ascribe and one cannot other-ascribe unless one can self-ascribe.

The Persons Theory of mentalizing is a no-priority symmetry account of the ascription of experiences. Strawson's account of self-consciousness is an interpersonal account of self-consciousness, meaning that self-consciousness must be understood as fundamentally related to consciousness of other persons, and vice versa, and this fundamental relation stresses the full symmetry between self-ascription and other-ascription. Since, the other views—TT and ST—do not stress full symmetry, then the Symmetry theory is a genuine alternative to those views.

Admittedly, this argument depends upon accepting the taxonomy and upon the idea that both TT and ST are asymmetrical accounts. To make a more clear and precise contrast between the Persons Theory and these views, I review the major tenets of TT and ST and illustrate how the Persons Theory differs from each (My taxonomy here draws from Goldman's, 2006, discussion of the similarities and differences between TT and ST.).

What are the major tenets of TT?

- First, mindreading depends upon the possession of a folk psychological theory. For example, in order for an individual to ascribe (whether to oneself

or to another) an experience of seeing a sheep, that individual has to possess a theory of seeing. The theory of seeing involves a variety of posits, for example material objects or sensations, that would describe, explain and predict what makes someone perceive something or other.

- Second, the folk psychological theory in question is analogous to a scientific theory, whether that theory is taken to be acquired or innate, in that it takes into account certain observable events and posits unobservable events to account for the perception.
- Third, the theoretical posits of the theory are causal laws that describe, explain and predict behavior. For instance, in the case of perception, the causal laws describe the phenomenon of seeing a sheep, including what it's like to experience it, explain how one sees a sheep, including the material, normative and physiological conditions, and predict what a perceiver seeing a sheep would do or what other perceivers given conditions would perceive.
- Fourth, the causal laws represent relations between mental states and non-mental events, between non-mental events and mental states, and between mental states and other mental states. For example, in the case of the ascription of a perception, the theory represents transitions from perceiving to acting, or from acting to perceiving or from perceiving to thinking.

The Persons Theory differs from TT in each of these four respects. First, Strawson is explicit that no theorizing about others is involved. We directly and immediately perceive the mental states of others, rather than infer from observed behavior to the posits that describe, explain and predict that behavior. Second, the

account of ascription in terms of persons is not analogous to a scientific theory, because it represents a truly commonsense or folk understanding of oneself and others as persons. To take another individual and oneself to be part of a *form of life* is not to present a scientific claim about that other individual and oneself, but, instead to experience the individual as adopting *the lived position of the species* (Cf. Avramides's, 2002, account of Strawson's notion of the person.). Third, the account of mindreading does describe, explain and predict, but it does so not by positing unobservable mental states, but instead by describing the embodied activities of human beings. This description of the activities of human beings can be so detailed that it serves the same explanatory and predictive purposes that the theory-theory does.

What are the major tenets of ST?

- First, mindreading does not depend solely on the possession of a folk psychological theory. (I say, “does not depend solely” because recently Alvin Goldman (2006), a major proponent of ST has proposed a hybrid account of mindreading in which theory is involved in thinking about others, although it is not central to the process itself. I will not go over the details of this view, however, as they are not central to the point.)
- Second, mindreading is a simulation process, or imaginative process, or an act of pretending that one is entertaining an experience. For example, if one is to ascribe the experience of seeing a sheep to an other person, then one pretends that one is that person seeing the sheep.
- Third, pretending that one is entertaining an experience realizes or implements the same mechanisms and processes as having that experience itself. For

example, the same process of seeing a sheep is mirrored in oneself as is involved in the individual actually seeing the sheep.

How does the Persons Theory differ from the simulation theory? Strawson is explicit that ascription of experiences to others is not like imagining or pretending, because one perceives that someone else is feeling depressed. Self-ascription and other-ascription, for Strawson are a kind of thoughtful perception. The mentalizing is perceptual in that the state of oneself and the other is grasped through a lived perceptual experience. The mentalizing involved in the Persons Theory is also thoughtful in that it involves recognition and acknowledgement of a person and that person's experiences. However, the view that Strawson suggests is not committed to any sort of mechanisms or processes of pretending, imagining or mirroring. The only mirroring that might be involved is a connection between the embodied personal properties of the other to which one is ascribing and the embodied personal properties of the individual engaging in the self-ascription. The claim is therefore that the infant develops the ability for the self-ascription of perceptions, actions and emotions, because it adopts a role in the game of persons—playing the game of perceiving, acting and emoting.

Other than the benefit of being a substantially new theory of mental ascription, is there any other support for the Persons Theory? Rejecting the theory-theory and the simulation theory accounts of mentalizing in favor of a third account is not new. Shaun Gallagher (2001) has argued that neither the theory-theory nor the simulation theory account for the import of “primary intersubjectivity” as a basic embodied skill and ability, and that each fails to capture the symmetrical relation between self and other that is basic to the phenomenology of lived experience. Both of these critiques can be read as

echoing what I have said here in favor of the Persons Theory. Gallagher and Dan Hutto (2008) have proposed a non-representational interaction theory of self and other through arguing that grasping of folk psychology is not a matter of grasping a theory (because causal explanation is too demanding for the infant) or pretending (because suppositional imagining is too demanding for the infant), but instead involves perception of bodily movements, gestures and facial expressions. I agree with both of these critiques of TT and ST.

Further, if you take the theory-theory and the simulation theory to be a matter of privileging the other and the self respectively, then it does not seem like there is much neuroscientific evidence for either prioritizing. If there were experimental evidence that self-ascription developed prior to other-ascription or that other-ascription developed prior to self-ascription, then such evidence could falsify the Persons Theory. However, MacRae et al. (2004) shows that both self-ascription (“I am seeing a sheep”) and other-ascription (“You are seeing a sheep.”) result in increased activity in the medial pre-frontal cortex. Maybe this is just because there is second-order or meta-cognition occurring in both cases, but in general this connection suffices to show that self-ascription and other-ascription are linked in the brain. A study might be developed to assess whether during joint engagements, but, prior to the ability to theorize and pretend, an infant is capable of other-ascription and self-ascription by this neuroscientific measure. If we discovered that the joint engagements involve the MPFC lighting up, then that might tell us how the joint engagements relate to the constitution of other-ascription and self-ascription. As far as finding some neuroscientific measure to falsify the Symmetry Thesis, in a review of the literature on the relation between self-knowledge and other-knowledge, Philip Robbins’

(forthcoming) survey of the neuroscientific evidence concerning priorities of self-ascription or other-ascription finds no evidence for such priorities. For example, Gillihan and Farrah (2005) argue that claims for the special status of self-related processing are premature because the paradigms that investigate self-ascription are not focused on a single system.

There is certainly more work to be done. One issue that remains to be discussed is how the infant and caregiver are governed by the rules of the game of joint engagement. How should we understand the rules of play that enter into the give and take that occurs during development? There is more work to be done on how embodied interactive play is rule- or norm-governed and how such rule- or norm-governance provides for a transition from primary intersubjectivity to secondary intersubjectivity. Another issue is how to make sense of the embodied gestures that enter into the recognition and acknowledgement. What exactly are the embodied activities that enable the mutual reciprocal relations to form between infant and caregiver? Another issue is that while I have discussed the import of shared experiences in joint engagement, I have not said much about the developmental move from such shared ascription of experiences to the individual ascription of experiences. (I am grateful to John Barresi for pressing this question.) It might be asked, “How does the infant and the early child come to ascribe experiences to herself independent of these joint engagements?” Because joint engagements are significant for the development of language, in particular for the development of the abilities to use the demonstratives ‘you,’ ‘I,’ and ‘we,’ narrative or linguistic competence plays an important role in advancing the developmental move from the kinds of self-ascription involved in joint engagement—what we might call ‘reflexive

self-ascription’—to the kinds of self-ascription involved in engagement with oneself in inner monologue or soliloquy—what we might call ‘reflective self-ascription.’ My concern in this paper has been with the former and more work needs to be done in considering how reflexive self-ascription contributes to reflective self-ascription. But, my hypothesis is that narrative or linguistic competence is central for this transition.

In closing, I have argued that the Persons Theory provides an argument for *the Symmetry Thesis*, which enables us to overcome the problem of other minds. What I am suggesting is that if we can return to our pre-Cartesian innocence—making sense of how self and other are on a par— then we can register that there is no real problem of the relation between ourselves and others. The argument for the Symmetry Thesis suggests that the awareness of ourselves and the awareness of others is on a par where priorities are concerned, namely that the self-other relation is a symmetrical relation, which obviates the need for an analogical inference from the self to other, or vice versa. The focus on joint engagement as an embodied and embedded practice enables us to see that the relation between self and other is a lived and practical relation. When we ask, “How is it possible for one to self-ascribe one’s experiences?”, the answer is that in order for one to self-ascribe one’s experiences it is *necessary for* one to ascribe an experiences to an other individual. The Persons Theory also gives us a new account of the nature of the joint engagements between individuals that is embodied and embedded in a physical and social environment, not a matter of a theoretical stance or a simulation routine.

Chapter 7

Dissolving the Conceptual Problem of Other Minds: Self-Ascription, Symmetry and Person Perception

In the introduction, I briefly discussed how the defense of the Symmetry Thesis might enable us to overcome the conceptual problem of other minds by focusing on the nature of self-consciousness. In the conclusion, I discuss the import of the Symmetry Thesis. This final chapter serves a summary of the dissertation. In addition, I hope to make connections with the philosophical problem of other minds. What I aim to present here are merely sketches of the problem, rather than a fully worked account. In section 1, I discuss the two distinct ways of setting up the problems of other minds: the Epistemological problem and the Conceptual problem. (My account of the problem of other minds is deeply indebted to Avramides (2001).) In section 2, I provide an outline of the approach to the epistemological problem and diagnose the problems with those approaches. In section 3, I outline of the approaches to the conceptual problem and diagnose the problems with those approaches. In section 4, I reiterate the Symmetry Thesis and the Persons Theory as a dissolution of the conceptual problem of other minds. In section 5, I respond to some final objections.

The problem of other minds is a distinctively *philosophical* problem. In everyday life, we ask questions like “How do I know what you are perceiving?” “How do I know your intentions?” or “How do I know what you are feeling?” But, in rare occasions only does this keep us from carrying out the activities of our everyday lives. The *philosophical* problem arises because of questions like “How do I know that there are any

minds like my own?” or “How do I know that other creatures have mental states, such as perceptions, intentions, emotions, thoughts, or experiences in general?”

In philosophy, the question about the problem of other minds is usually discussed as an epistemological problem. *The Epistemological Problem* is a problem of how we are justified in believing that others have minds or that others have experiences like ours. The assumption that forces us into the problem is that our beliefs about our own minds are justified in a *direct, immediate and transparent way*, but we can only justify beliefs about other minds in an *indirect, mediated and opaque way*. But, is the assumption that we are justified in our beliefs about our own minds in this *direct, immediate and transparent way* impervious to challenge?

The assumption that we KNOW our own minds in this direct, immediate and transparent way assumes that we *innately* possess the concepts “mind”, “experience”, “perception”, “action”, “emotion” and other mental concepts.

It assumes that all the concepts that enter into our consciousness of our selves and our experiences are given. *The Conceptual Problem* is a problem of how we are able to apply the concepts “mind”, “experience”, “perception”, “action”, “emotion” and other mental concepts to other individuals. But, is the assumption that we *innately* possess the concepts that enter into self-consciousness impervious to challenge?

As I discussed in the introduction, these are distinctively Cartesian assumptions about the distinctions between privacy and publicity and a claim about the link between the privacy and publicity. The privacy assumption might be understood to have two versions: the knowledge version and the concept version. The knowledge version is “I know my own mind in a direct and immediate way.” The concept version is “The

concepts of “mind” and mental states, e.g., “experience”, “perception” and other concepts can apply to myself privately.” The publicity assumption also might be understood to have two versions: the knowledge version and the concept version. The knowledge version of the publicity assumption is “I do not know others’ minds in a direct and immediate way; and if I do know other minds, only problematically.” The concept version of the publicity assumption is “The concepts of “mind” and mental states, e.g., “experience”, “perception” and other mental concepts may not apply to others; and if they do, only problematically.” The last assumption of the argument that generates the problem of other minds is an assumption about the link between privacy and publicity. There is an assumption about that there is no logical or conceptual link between our knowledge of or our concepts of our own minds, on the one hand and our knowledge of or our concepts of others’ minds, on the other hand.

This provides us with a chasm between our awareness of our own minds and our awareness of other minds and gives rise to an Epistemological Problem: *How is it possible to have knowledge of other minds?* One helpful way to talk about this problem might be to consider Thomas Nagel’s discussion of skepticism. Nagel (1986, p. 68) discusses three different answers to skepticism. The first is what is a kind of skeptical forfeit, giving up on bringing the chasm between self and other. The second is a kind of reductive answer that doesn’t give up on bridging the gap, but translates the problem into a problem that can be solved, but ultimately does not meet the initial problem. The Heroic Solution is to build a tailored theory to overcome the problem. But, each of these accounts presuppose the problem of other minds is an epistemological problem. However, as I discussed above, the epistemological problem is ultimately based in a

conceptual problem: *How is it possible to possess concepts of other minds?* In that case, it deserves a conceptual dissolution, rather than any of these other answers. I will now turn to the reasons why the other solutions are problematic.

The skeptical forfeit involves accepting the assumptions that generate the problem and forfeiting any attempt to bridge the chasm between the self and other. One historical position is solipsism, which forfeits under the force of the problem of other minds (says, “Uncle!!”) and asserts that “I am the only mind that exists” or “My experiences are the only experiences that exist.” However, solipsism is an incoherent doctrine, for in order for the solipsist to say and think the claims that express the view, other language users and other minds must exist.

A reductive answer to the epistemological problem of other minds interprets the question “How are we justified in believing that there are other minds?” as if the question were asking less. The internal reductive answer argues that the resources to make a connection between self and other can be found internal or within the self. This answer to the problem accepts the publicity assumption— I do not know others’ minds in a direct and immediate way— but suggests that since we know our own minds in the past (memory) and in the future (projection), this is sufficient to address the problem. The internal reductivist says, “An other within my own mind is an other enough for me...” The external reductive answer argues that the resources to make a connection between self and other can be found external or without the self. This answer to the problem denies the privacy assumption— I know my own mind in a direct and immediate way— but suggests that since we know BOTH our own and others’ minds in an indirect and mediated way, this is sufficient to address the problem. The external reductivist says,

“The self within my own mind is nothing for me...” However, reductive answers fail to take the problem of other minds seriously, because the reductive answer reinterprets the challenge as if it were not a challenge at all.

The Heroic solution acknowledges the gap between self and other in a way that takes the problem to be a serious challenge and tries to leap across the gap without narrowing it. About heroic attempts to overcome skepticism about the external world, Nagel says “The chasm below is littered with epistemological corpses” (Nagel (1986, p. 69)). Heroic solutions include Plato’s theory of the forms and Descartes argument for the connection between mind and world. What is the heroic solution to the problem of other minds? The traditional heroic solution uses the argument from analogy to attempt to justify belief in other minds. The argument from analogy is the following:

1. I am aware of my own mental states in a direct and immediate way.
2. There are correlations between my own mental states and my behavior and bodily states.
3. I am aware of the behavior and bodily states of others.
4. There are correlations between my own behavior and bodily states and the behavior and bodily states of others.
5. Therefore, the best explanation of these sets of correlations is that others have minds and mental states.

The two main objections to the argument from analogy is the one-case objection and the correlations objection. The One-Case Objection contends that the argument from analogy proceeds from the direct and immediate knowledge of one’s own experience—for instance, the awareness of a pain—to the indirect and mediated knowledge of others’

bodily states and behavior. However, inference cannot be based on merely one case, for that would render the inference invalid (Cf. Malcolm's (1962)). The Correlations Objection contents that the argument from analogy assumes that there are two sets of correlations between the elements of (a) my mental states and my behavior and bodily states and (b) my behavior and bodily states and the behavior and bodily states of others.

However, there are occasions in which the correlations between the elements of (a) and (b) simply do not hold (Ryle (1949)); and, even if they did hold, correlation is not sufficient for causation (Cf. Hyslop and Jackson (1972)).

The alternative, therefore, is to dissolve the epistemological problem of other minds by diagnosing the problem with the argument from analogy and proposing a new strategy. I begin with articulating the structure of analogy in general, as this is instructive for understanding what has gone wrong with the argument from analogy. The first step in an analogy is to find a target or purpose of using the analogy, for example, "What do we need to see, do or understand?" In the case of the argument from analogy, the target is making sense of how we have knowledge of or thought about other minds. The second step in an analogy is to begin with a source, for example, "What do we already see, do and understand that might help?" In this case, the source is our knowledge of or thought about our own minds. The third step is mapping the source to the target through finding correspondences between source and target. In this case, we map correspondences between awareness of our own minds and awareness of other minds. The last step is to adapt through a reflective equilibrium the mapping of source and target so that our analogy achieves its purpose. I propose that the failure of the argument from analogy be diagnosed as a failure of one or other of these steps. In particular, the failure of the

mapping step to capture the symmetrical relation between self and other. I will say more about this below.

For now, I want to engage in some recent philosophical history. On the American side of the Atlantic, in the 1970's there was little concern with the problem of other minds. In America since the 1970's because of the rise of cognitive science, focus on the epistemological problem of other minds began to fade. As a quote from Jerry Fodor suggests: "When I was a boy in graduate school, the philosophy of mind had two main divisions: the mind/body problem and the problem of other minds... Philosophical fashions change. It's gotten hard to believe that there is a *special* problem about the knowledge of other minds (as opposed to knowledge of anything else)" (1994). As I discussed in the introduction, however, Jerry Fodor's methodological solipsism is a form of Cartesian individualism. In the late 20th century, in America, the epistemological problem was transmuted into an investigation in philosophical psychology in the theory of mind. Philosophers and psychologists were not focused on a distinctively *philosophical* problem (although, it remained in the background), but instead, wanted to inquire into the nature and origin— both evolutionary (phylogenetic) and developmental (ontogenetic)— of our ability to attribute mental states to others, namely how our human ancestors and human infants come to possess a theory of mind. What did they want to describe, explain and predict?

For the most part, theorists have been interested in accounting for what I discussed in the last chapter as the intersubjective triangle. There are roughly two accounts of mental state attribution or 'mentalizing': the theory theory and the simulation theory. While in the last chapter I discussed the major tenets of each theory and

presented an argument for the claim that the Persons Theory differs from these two theories, I will now provide some objections to each of the theory theory and the simulation theory in order to provide further impetus for the Persons Theory. To review, according to the theory theory, the ascription of experiences is made possible by the possession of a theory of other minds. According to the simulation theory, the ascription of experiences is made possible by the simulation of other minds.

The theory-theory has gained support because it accounts for our ascription of mental states to others in third-personal terms, which are assumed to be less problematic than first-personal or second-personal terms. Theorizing about other minds is continuous with theorizing about the world of objects, properties and relations. The theory-theory has also gained empirical support from evidence that infants below the age of four fail the false-belief test. However, there are some clear objections to the use of the theory-theory. The first objection is that the theory-theory supposes that our thought about other minds should be understood predominantly in terms of beliefs, desires and other intentional states (what we might call ‘the mentalistic supposition’), but this is not the way our interaction with others is phenomenologically (Gallagher (2001, p. 94)). The theory-theory also supposes that infants have the capacity to provide causal explanations and predictions (what we might call ‘the causal explanation supposition’), but while they may have causal understanding, they do not develop the ability to provide causal explanations and predictions until much later (Roessler (2002, p. 236)). Each of these objections suggest that the Theory-Theory does not account for the import of “primary intersubjectivity” as a basic embodied skill and ability, and fails to capture the symmetrical relation between self and other that is basic to the phenomenology of lived

experience. However, as I argued in the last chapter, the Persons Theory provides a different account, because our ascription of experiences is based in perception of others and ourselves rather than predominantly thought about others and ourselves. Also, the Persons Theory rejects the causal explanation assumption.

The Simulation Theory has gained support because it accounts for our attribution of mental states to other individuals in first-personal terms, which are assumed to be less problematic than third-personal or second-personal terms. According to the Simulation Theory, theorizing about other minds is continuous with theorizing about ourselves, including our own minds and our own experiences. Simulation Theorists (Goldman (2006)) have argued that there is empirical support for the Simulation Theory from evidence of the existence of mirror neurons. However, the Simulation Theory supposes that our thought about other minds should be understood predominantly in terms of beliefs, desires and other intentional states (mentalistic supposition), but this is not the way our interaction with others is phenomenologically (Gallagher (2001: 94)). The Simulation Theory supposes that infants have the capacity to simulate— through imaginative projection— the mental states of others (the imaginative projection supposition), but while they may have the capacity for imagination, they do not develop the ability to engage in imaginative projection until much later (Roessler (2001: 236)).

Each of these objections suggest that the simulation theory does not account for the import of “primary intersubjectivity” as a basic embodied skill and ability, and fails to capture the symmetrical relation between self and other that is basic to our phenomenology of lived experience.

In the last chapter, I also discussed how each of the dominant theories— the theory theory and the simulation theory— are taxonomized with respect to self-other priority. The Theory-Theory is an *other-priority asymmetry view* because the ability to other-ascribe is not dependent upon the ability to self-ascribe in the sense that the ability to other-ascribe is prior to the ability to self-ascribe. The Simulation Theory is a *self-priority asymmetry view*: the ability to self-ascribe is not dependent upon the ability to other-ascribe in the sense that one need not think about or perceive others in order to self-ascribe. However, notice that the objections to both the Theory-Theory and the Simulation Theory arise because both theories are committed to claims that have an analogical structure, which I take to be the major downfall for each theory in allowing us to overcome the conceptual problem of other minds. On the theory-theory, the ability to self-ascribe is analogically based on the ability to other-ascribe, since one theorizes about others' mental states first, and then theorizes about one's own mental states. On the simulation theory, the ability to other-ascribe is analogically based on the ability to self-ascribe, since one experiences one's own mental states first, and then simulates an other's mental states.

But, is there any alternative? Can we escape from claims that depend upon an analogical structure? If we could, would that enable us to dissolve the conceptual problem of other minds? As the epistemological problem of other minds fades and philosophers and psychologists (on both sides of the pond) fight the war between the Theory-Theory and the Simulation Theory, British Analytic philosophers inspired by Kant are developing an approach to the conceptual problem of other minds, asking the distinctive question that I raised in the introduction and Chapter 1, "*How is it possible* for

us to ascribe mental states to ourselves and to others?” Again, Peter Strawson concerned with the conceptual problem, writes, “it is a necessary condition of one’s ascribing states of consciousness, experiences, to oneself, in the way that one does, that one should also ascribe them, or be prepared to ascribe them, to others who are not oneself” (1959, p. 94); and “I am inclined to think that at least some [transcendental arguments are strictly valid] (e.g., self-ascription implies the capacity for other-ascription)...” (1985, p. 25). Gareth Evans, writes, “it is essential that I am aware that the person of whom I am so thinking *is myself*; certainly I must have in mind what it is for [δ is dead] to be true, for arbitrary δ , but I must also have in mind what it is for [$\delta=I$] to be true. My thought about myself does satisfy the Generality Constraint; and this is because I can make sense of identifying a person, conceived from the standpoint of an objective view of the world, as myself” (1982, p. 210). I take these two thinkers to be answering the question “*How is it possible for one to think about an other subject’s experience, for instance an other subject’s experience of seeing a sheep?*” and “*How is it possible for an other to think about one’s own experience, for instance for an other subject to think of my experience of seeing a sheep?*”

According to the proponents of the British Neo-Kantian Invasion, the best way to approach a how-possible-X question is to answer by showing what is necessary for X, which means that a transcendental argument is required. In chapter 2, I argued that a transcendental argument was necessary for this task. For instance, we ask, “How is it possible for one to self-ascribe one’s experiences of seeing a sheep?” and one answer might be that in order for one to self-ascribe one’s experience of seeing a sheep it is necessary for one to ascribe an experience of seeing a sheep to an other individual. I

suggested then that one way to approach the conceptual problem of other minds is to argue for a thesis that argues that consciousness of one's own mind is constitutively linked to consciousness of an other subject's mind. Throughout the dissertation, I followed Bermúdez (1998: Chapter 9) in thinking that the problem of other minds can be addressed by arguing for *the Symmetry Thesis* about self-ascription. Again *the Symmetry Thesis (ST)*: “A subject's psychological self-awareness [self-ascription] is constitutively linked to his awareness of other minds [other-ascription]” (230); *In order for one to have the ability to self-ascribe experiences, one must have the ability to ascribe experiences to other beings like oneself.*

Remember that in the argument from analogy, in the Theory-Theory, and in the Simulation Theory, there is *an analogical structure* in the relationships between self-ascriptions and other-ascriptions. As Bermúdez points out, “on most versions of the argument from analogy, there is no constitutive connection between first-person and third-person ascriptions of psychological properties” (1998, p. 231). Throughout the dissertation, I have argued that self-ascription and other-ascription are symmetrical? My claim is that the argument for the Symmetry Thesis enables us to obviate the analogical structure and that an adequate defense of the Persons Theory will enable us to overcome the conceptual problem of other minds. I say ‘overcome,’ because the strategy involves more of a dissolution of the problem than a solution to the problem. I have argued that if we can make sense of how self-ascription and other-ascription are mutually reciprocally dependent and find evidence of such a dependence, then we can dissolve the conceptual problem of other minds. As Ludwig Wittgenstein has said, “the real discovery is the one that makes me capable of stopping doing philosophy when I want to. – The one that gives

philosophy peace, so that it is no longer tormented by the questions that bring *itself* into question. – Instead, we now demonstrate a method, by examples; and the series of examples can be broken off. – Problems are solved (difficulties are eliminated), not a *single* problem. There is not a philosophical method, though there are indeed methods, like different therapies” (Wittgenstein, 1953, §133). And, if the Persons Theory is correct, then we might find a therapy by considering the phenomenon of joint engagement.

As Gallagher has argued against the theory theory and simulation theory, “...understanding others in everyday life does not usually involve either taking a theoretical stance or deploying a simulation routine. It depends instead on a capacity for embodied practice that begins early (and is likely to be partially innate) and continues throughout normal (non-pathological) experience. Thus, in contrast to the strong pragmatic claim for theory of mind, namely that it is our primary and pervasive means for understanding other persons, I would substitute a strong pragmatic claim for primary intersubjectivity. It is not just primary in developmental terms: it continues to characterize most of our interpersonal interactions, and it forms the basis for the more specialized mentalistic interpretations of how others perform in the practice of their own minds.” (2001, p. 104)

The nature of the joint engagements between individuals is embodied and embedded in a physical and social environment, not a matter of a theoretical stance or a simulation routine. What I am suggesting is that if we can return to our pre-Cartesian innocence, then we can register that there is no real problem of the relation between ourselves and others. The constructive argument for the Symmetry Thesis that I

presented in the dissertation suggests that the awareness of ourselves and the awareness of others is on a par where priorities are concerned, namely that the self-other relation is a symmetrical relation, which obviates the need for an analogical inference from the self to other, or vice versa.

The focus on joint engagement as an embodied and embedded practice enables us to see that the relation between self and other is not a theoretical stance or a simulation process, but instead, is a lived, practical and everyday relation. When we ask “How is it possible for one to self-ascribe one’s experiences?” the answer is that in order for one to self-ascribe one’s experiences it is NECESSARY FOR one to ascribe an experiences to an other individual. I will now consider some possible objections.

The first objection might be that I am not taking the conceptual problem of other minds seriously, because I have just ignored the problem. According to this objection, I have just asserted that we are embedded in the lived position of being a person among persons. To return to the metaphor I used above, rather than using a heroic solution, the objection insists, I have just said we are already on other side of the chasm without realizing it. However, I have provided a transcendental argument that self-ascription depends on other-ascription and have shown that our default status is being a person among persons, in the sense that developmental evidence suggests that our natural position and the intersubjective position.

The second objection might be that if we accept *the Symmetry Thesis* and Embodied Interaction Theory, then because self-ascription and other-ascription are constituted in tandem, then: (P1) We cannot account for how we have a private awareness of ourselves (privacy objection); (P2) We cannot account for how we have a

public awareness of others (publicity objection). My response to the privacy objection is that the experience of what developmental psychologists call “the private self” does not arise until late in development and does not arise unless there are cultural memes available like talk of souls or minds *qua* private beings. So, it is an expendable idea and not constitutive of self-consciousness. My response to the publicity objection is that while according to the Persons Theory, we do not ascribe experiences by way of theories or simulations of others, our bodies and behavior and the others’ bodies and behavior are public. I will now provide some illustration of the kind of research that might be done concerning *the Symmetry Thesis* and the Persons Theory by considering another objection.

The Symmetry Thesis says *In order for one to have the ability to self-ascribe experiences, one must have the ability to ascribe experiences to other beings like oneself.* But, the objector suggests, “This implies that feral children (children raised by non-human animals) would be aware of themselves as non-human animals and domesticated non-human animals (horses, dogs and cats) would be aware of themselves as humans, but this is clearly absurd.” My reply is that these concerns are at best speculative, but the implication is not necessarily a downfall of the account, because it helps us to understand why *Bildung*, or social formation is pivotal. When we discover feral children, getting them to rejoin human society is incredibly difficult; when we train non-human animals, we are encouraging them to join human society in some respects.

One might also object that I have not provided enough empirical evidence for the Persons Theory. But, there are two possible lines of evidence in psychology for the Symmetry Thesis that I have not discussed. If self-ascription requires other-ascription,

then there should be evidence that where there is little other-ascription, there are failures of self-ascription. As I discussed briefly in Chapter 6, this prediction is found true in multiple cases of the symptoms of autism. Also, if other-ascription constitutes self-ascription, then there should be evidence that where there is an abnormal increase in other-ascription, there is an abnormal increase in self-ascription. This prediction is found true in multiple cases of symptoms of schizophrenia. But, both of these claims are merely predictions about the import of the Symmetry Thesis and the Persons Theory.

Of course, more work needs to be done. But, I have argued that the argument for the Symmetry Thesis and the defense of the Persons Theory provides us with a way to dissolve the conceptual problem of other minds. This enables us to see that when I think I am experiencing a sheep, I am not alone in my self-consciousness. To think that I am experiencing a such and such is to think that we are experiencing a such and such. The I is a We and the We is an I. And, the We is not composed of two self-standing subjectivities, but of one, which has its dualism at a non-psychological level. Each of us is alone by being distinguished bodily, but whenever we are conscious of our experiences the other is always present, namely the I needs an Other.

This is central for the infant because the infant becomes an individual that is capable of self-ascribing perceptions under the structure of the person in and through interaction with other creatures like itself. It is in that sense that joint engagement is constitutive of self-ascription of experiences, because it is not only temporally and causally prior to self-ascription of experiences, but the structure under which one self-ascribes provides for the nature and essence of self-ascription. To think “I am experiencing the world” is to think “We are experiencing the world.”

Bibliography

Anscombe, E. (1994). The First Person. In *Self-Knowledge*. Ed. Quassim Cassam.

New York, NY: Oxford University Publishing.

Armstrong, D. (1981). *The Nature of Mind and Other Essays*, Ithaca, NY: Cornell

University Press

Austin, J. L. (1939). Are There A Priori Concepts? *Proceedings of the Aristotelian Society*. 18.

Avramides, A. (2001). *Other Minds*. New York, NY: Routledge.

Ayers, M. (1991). *Locke*. 2 vols. New York, NY : Routledge.

Baker, M. (2003). *Lexical Categories*. Cambridge, UK: Cambridge University Press.

Bar-On (2004) *Speaking My Mind*. New York, NY: Oxford University Press

Bar-On, D. (2004). *Speaking My Mind*. New York, NY: Oxford University Press.

Barresi, J. (1999). On Becoming a Person. *Philosophical Psychology*, 12, 79–98.

Barresi, J. (2004). Intentional Relations and Divergent Perspectives in Social

Understanding. In S. Gallagher and S. Watson (Eds.) *Ipseity and Alterity:*

Interdisciplinary Approaches to Intersubjectivity. (pp. 74–99). Rouen: Presses

Universitaires de Rouen

Barresi, J. (Forthcoming). On Seeing Our Selves and Others as Persons. *New Ideas in*

Psychology.

Barresi, J. and Moore, C. (1996). Intentional Relations and Social Understanding.

Behavior and Brain Sciences. 19(1), 107–122.

- Barresi, J. and Moore, C. (2008). The neuroscience of social understanding. In J. Zlatev, T. Racine, C. Sinha and E. Itkonen (Eds.) *The Shared Mind: Perspectives on Intersubjectivity*, Amsterdam/Philadelphia: John Benjamins.
- Beiser, F. (2002). *German Idealism*. Cambridge, MA: Harvard University Press.
- Bermúdez, J. L. (1995). Ecological Perception and the Notion of a Non-Conceptual Point of View. In J. L. Bermúdez, A. Marcel and N. Eilan (Eds.), *The Body and the Self* (pp. 153–174). Cambridge, MA: MIT Press.
- Bermúdez, J. L. (1998). *The Paradox of Self-Consciousness* Cambridge: MIT Press.
- Bermúdez, J. L. (2005). *Philosophy of Psychology*. New York, NY: Routledge.
- Bermúdez, J. L., Marcel, A., and Eilan, N. (1995). *The Body and the Self*. Cambridge, MA: MIT Press
- Bertenthal, B. I., Proffitt, D. R., Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *Journal of Experimental Child Psychology* 37, 213–30.
- Bogdan, R. J. (2000). *Minding Minds*. Cambridge, UK: MIT Press.
- Bonatti, L. et al. (2002). The human first hypothesis: identification of conspecifics and individuation of objects in the young infant. *Cognitive Psychology* 44, 388–426.
- Brandom, R. (1994). *Making it Explicit*. Cambridge, MA: Harvard University Press.
- Brinck, I. (1999). Self-Identification and Self-Reference *Electronic Journal of Analytic Philosophy*, May. Special issue on the philosophy of Gareth Evans. ed. Rick Grush. <http://www.phil.indiana.edu/ejap/ejap.html>
- Brook, A. (1999). Circularity or Reduction?: What is the real issue? Online: http://host.uniroma3.it/progetti/kant/field/bermudezsymp_brook.htm

- Brook, A. (1994). *Kant and the Mind*. Cambridge, UK: Cambridge University Press.
- Brook, A. (1997). Self and World, by Quassim Cassam. *Mind, New Series*. 110.437: 190–196.
- Brook, A. and Devidi, R. (2001). *Self-Reference and Self-Awareness*. Amsterdam, Netherlands: John Benjamins Publishing Co.
- Brooks, R. & Meltzoff, A.N. (2002). The importance of eyes: how infants interpret adult looking behavior. *Developmental Psychology*, 38, 958–966.
- Brueckner, A. (1996). Modest Transcendental Arguments *Philosophical Perspectives* 10, pp. 265–66.
- Brueckner, A. (2005). Review of Speaking My Mind.
- Burge, T. (1979). Individualism and the Mental *Journal of Philosophy* 85 (November):649–63.
- Butler, J. (1736). Of Personal Identity In *Personal Identity*. Ed. John Perry. Los Angeles, CA: University of California Press.
- Butterworth G. and Jarrett N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*. 1991;9: 55–72.
- Campbell, J. (1994). *Past, Space and Self*. Cambridge, MA: MIT Press.
- Campbell, J. (1999a). Schizophrenia, the space of reasons and thinking as a motor process. *The Monist*, 82 (4): 609–625.
- Campbell, J. (1999b). Immunity to error through misidentification and the meaning of a referring term. *Philosophical Topics*, 26: 89–104.

- Campbell, J. (2005). Joint Attention and Common Knowledge. In Eilan, N., Hoerl, C., McCormack, and Roessler, J. (Eds). *Joint Attention: Communication and Other Minds*. (pp. 287–297). Oxford, UK: Oxford University Press.
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63 (4, Serial No. 255).
- Cassam, Q. (1994). *Self-Knowledge*. New York, NY: Oxford University Press
- Cassam, Q. (2007). *The Possibility of Knowledge*. New York, NY: Oxford University Press.
- Cassam, Q. (1987). Transcendental Arguments, Transcendental Synthesis, and Transcendental Idealism *Philosophical Quarterly* 37
- Cassam, Q. (1995). Introspection and Bodily Self-Ascription. *The Body and the Self*. Ed. Bermudez, Marcel, and Eilan Cambridge, MA: MIT Press
- Cassam, Q. (1997). *Self and World*. New York, NY: Oxford University Press
- Cassirer, E. (1927). *The Philosophy of Symbolic Forms*. Trans. Ralph Manheim (1955) New Haven, CT: Yale University Press.
- Casteneda, H. (1999). *The Phenomenologic of the I*. Bloomington, ID: Indiana University Press
- Cavell, S. (2002). *Must We Mean What Say?* Cambridge: Cambridge University Press
- Chalmers, D. (1996). *The Conscious Mind*. New York, NY: Oxford University Press.
- Chisholm, R. (1981). *The First Person*. Minneapolis: University of Minnesota Press.
- Chisholm, R. (1977). *Theory of Knowledge*, 2nd ed., Englewood Cliffs: Prentice Hall.
- Crane, T. (2001). *History of the Mind-Body Problem*. London, UK: Routledge.

- Davidson, D. (1996) Subject, Intersubjective, Objective in P. Coates and D. Hutto (eds) *Current Issues in Idealism* Thoemmes Press
- Davies, M. (1994). The Mental Simulation Debate. *Philosophical Issues* 5:189-218.
- De Gaynesford, M. (2003). "Is I Guaranteed to Refer?" *Pacific Philosophical Quarterly*. 84: 109–26.
- Dennett, D. (1991). *Consciousness Explained*. Boston, MA: Little Brown and Co.
- Descartes, R. (1641/1984). 'Meditations on First Philosophy', in Cottingham, Stoothoff and Murdoch (trans.), *The Philosophical Writings of Descartes Volume II*, Cambridge: Cambridge University Press.
- Eilan, N. (2002). Joint Attention, Communication and Mind In Eilan, N., Hoerl, C., McCormack, and Roessler, J. (Eds). *Joint Attention: Communication and Other Minds*. (pp. 1–33). Oxford, UK: Oxford University Press.
- Eilan, N. and Roessler, J. (2003). *Agency and Self-Awareness*. New York, NY: Oxford University Press.
- Eilan, N., Hoerl, C., McCormack, T., and Roessler, J. (2005). *Joint Attention: Communication and Other Minds*. New York: NY; Oxford University Publishing.
- Evans, G. (1982). *The Varieties of Reference*. John McDowell (Ed.) Oxford: Oxford
- Fichte, J. G. (2000). *Foundations of Natural Right*. Trans. Michael Baur. Ed. F. Neuhauser. Cambridge, UK: Cambridge University Press.
- Fodor, J. (1980). Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology In *The Nature of Mind*. Ed. David Rosenthal. New York, NY: Oxford University Publishing. pp. 485–498

- Fodor, J. (1980). Author's Response In *The Nature of Mind*. Ed. David Rosenthal. New York, NY: Oxford University Publishing. pp. 500–501
- Fodor, J. (1994). Jerry Fodor In *A Companion to the Philosophy of Mind* Malden, MA: Blackwell Publishers
- Franks, P. (2005). Post-Kantian Transcendental Arguments In *All or Nothing: Systematicity, Transcendental Arguments and Skepticism in German Idealism*. Ed. Paul Franks. Cambridge, MA: Harvard University Press.
- Frege, G. (1956). The Thought *Mind* 65.259: 289–311.
- Fricker, Elisabeth (1998). Self-Knowledge: Special Access Versus an Artifact of Grammar— A Dichotomy Rejected. *Knowing Our Own Minds*. Ed. Wright, Smith, and MacDonald New York, NY: Oxford University Press
- Frith, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gallagher, S. (2000). Self-reference and schizophrenia: A cognitive model of immunity to error through misidentification. In D. Zahavi (ed). *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-experience* (pp. 203-39) Amsterdam: John Benjamins.
- Gallagher, S. (2001). The practice of mind: Theory, simulation, or primary interaction. *Journal of Consciousness Studies*, 8 (5-7): 83–108.
- Gallagher, S. (2004). Neurocognitive models of schizophrenia: A neurophenomenological critique. *Psychopathology* 37: 8–19. Invited paper with response by Christopher Frith. Comments on Shaun Gallagher. *Psychopathology*, 37 (2004): 20–22

- Gallagher, S. (2005). *How the Body Shapes the Mind*. New York, NY: Oxford University Press.
- Gallagher, S. and Hutto, D. (2008). Understanding others through primary interaction and narrative practice. In Zlatev, Racine, Sinha and Itkonen (eds). *The Shared Mind: Perspectives on Intersubjectivity* (pp. 17–38). Amsterdam: John Benjamins
- Gallagher, S. and Shear, J. (1999). *Models of the Self*. Thorverton, UK: Imprint Academic
- Gallagher, Shaun and Meltzoff, Andrew. (1996). The Earliest Sense of Self and Others: Merleau-Ponty and Recent Developmental Studies. *Philosophical Psychology*. 9: 213–236.
- Gallup, G. G., Jr. (1970). Chimpanzees: self recognition. *Science*, 167, 86–87.
- Gertler, B. (2008). Self-Knowledge *Stanford Encyclopedia of Philosophy* Online: <http://plato.stanford.edu/entries/self-knowledge/>
- Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Gillihan, S. J. and Farah, M. J. (2005). Is Self Special? A Critical review of evidence from experimental psychology and cognitive neuroscience. *Psychology Bulletin* Jan, 131.1: 76–97.
- Goldman, A. (2006) *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading*. Oxford, UK: Oxford University Press
- Grene, M. (2002). *The Philosophy of Marjorie Grene* edited by Randall E. Auxier and Lewis Edwin Hahn

- Grush, R. (2001). Self, world and space: on the meaning and mechanisms of egocentric and allocentric spatial representation. *Brain and Mind* 1(1): 59–92.
- Grush, R. (2002). Cognitive Science. In Machamer and Silberstein (Eds.) *Guide to Philosophy of Science*. Oxford, UK: Basil Blackwell.
- Gunther, Y. (2003). *Essays on Nonconceptual Content*. Edited. Cambridge, MA: MIT Press.
- Hacker, P. M. S. (2001). Strawson's Concept of Being a Person. *Proceedings of the Aristotelian Society*, New Series, Vol. 102, (2002), 21–40.
- Haugeland, J. (1982). Heidegger on Being a Person *Noûs*, Vol. 16, No. 1, 1982 A. P. A. Western Division Meetings (Mar., 1982), 15–26.
- Heal, J. (1994). Simulation Vs. Theory-Theory: What is at Issue? In Peacocke, C. (Ed.) *Objectivity, Simulation and the Unity of Consciousness*. (pp. 129–144). Oxford, UK: Oxford University Press.
- Hegel, G. W. F. (1977). *Phenomenology of Spirit*. Trans. A. V. Miller Oxford, UK: Oxford University Press.
- Heidegger, M. (1997). *Kant and the Problem of Metaphysics*. Trans. Richard Taft. Bloomington, IN: Indiana University Press.
- Henrich, Dieter (1971). Fichte's Original Insight in *Contemporary German Philosophy vol. I*, trans. David R. Lachterman. University Park: Pennsylvania State Univ. Press, 15–53.
- Hobson, P. (2002). *The Cradle of Thought*. Oxford, UK: Oxford University Press.

- Hobson, P. (2005). What Puts the Jointness in Joint Attention? In Eilan, N., Hoerl, C., McCormack, and Roessler, J. (Eds). *Joint Attention: Communication and Other Minds*. (pp. 185–204). Oxford, UK: Oxford University Press.
- Hofstadter, D. (2007). *I am a Strange Loop*. New York, NY: Basic Books.
- Honneth, A. (1996). *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. Cambridge, UK: Polity Press.
- Hornsby, J. (1980). *Actions*. London: Routledge and Kegan Paul.
- Hubley, P. & , C. Trevarthen. (1979). Sharing a task in infancy. In Uzgiris, I. C. (Ed.) *Social Interaction During Infancy: New Directions in Child Development, Vol. 4* (pp. 57–80). San Francisco: Jossey-Bass.
- Hume, D. (1978). *A Treatise of Human Nature*. Ed. Selby-Bigge and Nidditch. Oxford, UK: Oxford University Press.
- Hurley, S. (1998). *Consciousness in Action*. Cambridge, MA: Harvard University Press
- Hyslop, A. and Jackson, F. C. (1972). The Analogical Inference to Other Minds
American Philosophical Quarterly, 9: 168–176
- Ishiguro, H. (1980). The Primitiveness of the Concept of the Person. In Zan Van Straaten (Ed.) *Philosophical Subjects: Essays presented to P. F. Strawson*.
- Ismael, J.T. (2007). *The Situated Self*. Oxford: Oxford University Press
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14, 201–211.
- Kagan, J. (1998). Is there a Self in Infancy? In *Self-Awareness*. New York: Guilford Press: 137–147.

- Kant, I. (1781). *Critique of Pure Reason*. Trans. Paul Guyer (1999). Cambridge, UK: Cambridge University Press.
- Korner, S. (1974). *Conceptual Frameworks*. Oxford, UK: Basil Blackwell.
- Kriegel, U. (2003). Consciousness as Intransitive Self-Consciousness: Two Views and an Argument. *Canadian Journal of Philosophy*, 33, 103–132
- Kriegel, U. (2008). Self-Consciousness. online: <http://www.utm.edu/research/iep/s/self-con.htm>
- Lehrer, K. (2002). Self-Representation, Representation and the Self. *Philosophy and Phenomenological Research* 64 (2): 412-430
- Lewis, M. and Carmody, D. (2008). Self-Representation and Brain Development *Developmental Psychology* 44.5: 1329–1334.
- Locke, J. (1689). *An Essay Concerning Human Understanding*. Indianapolis, IN: Hackett Publishing
- Ludlow, P. and Martin, N. (1998). *Externalism and Self-Knowledge*. Stanford, CA: CSLI Publications
- Lycan, W. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.
- Mackie, J. L. (1963). Are there any Incorrigible Empirical Statements? *Australasian Journal of Philosophy* 41 (May):12–28.
- Mackie, J. L. (1976). *Problems from Locke*. Oxford, UK: Clarendon Press.
- Macrae, C., T. Heatherton, and W. Kelley (2004). A Self Less Ordinary: The Medial Prefrontal Cortex and You In *The Cognitive Neurosciences III*. Cambridge, MA: MIT Press: 1067–1076.
- Malcolm, N. (1962). Knowledge of Other Minds In V. C. Chappell (Ed.) *The Philosophy*

- of Mind* Englewood Cliffs, NJ: Prentice Hall.
- Marcel, A. (2003). The Sense of Agency: Awareness and Ownership of Action. In *Agency and Self-Awareness*. Ed. Roessler and Eilan Oxford, UK: Oxford University Press
- McDowell, J. (1996). Mind and World Cambridge, MA: Harvard University Press
- McDowell, J. (1998a). Response to Crispin Wright In *Knowing Our Own Minds*. Ed. Crispin Wright, Barry C. Smith, and Cynthia Macdonald. New York, NY: Oxford University Press, pp. 47–62.
- McDowell, J. (1998b). Referring to Oneself. *The Philosophy of P. F. Strawson*. Ed. Hahn Chicago, IL: Open Court
- Mead, G. H. (1934). *Mind, Self, and Society*. Chicago, IL: University of Chicago Press.
- Meeks, R. (Diss). *Identifying the First Person*.
- Meltzoff, A. N. (1988). Infant imitation and memory: Nine-month olds in immediate and deferred tests. *Child Development*, 59, 217–225.
- Meltzoff, A. N. (2007). The ‘like me’ framework for recognizing and becoming an intentional agent. *Acta Psychologica*, 124 26–43.
- Meltzoff, A.N., & Moore, M.K. (1994). Imitation, memory, and the representation of persons. *Infant Behavior and Development*, 17, 83–99.
- Merleau-Ponty, M. (1981). *The Phenomenology of Perception*. Trans. Colin Smith. New York, NY: Routledge
- Metzinger, T. (2003). *Being No One*. Cambridge, MA: MIT Press.
- Metzinger, T. (2009). *The Ego Tunnel*. New York, NY: Basic Books.

- Moore, C. and Dunham, P. (1995). *Joint Attention: Its Origins and Role in Development*. New Jersey Hove, UK: Lawrence Erlbaum Associates, Publishers.
- Moran, R. (2001). *Authority and Estrangement*. Princeton, NJ: Princeton University Press.
- Moran, Richard (2001). *Authority and Estrangement*. Princeton, NJ: Princeton University Press
- Mulhall, S. (2007). *Wittgenstein's Private Language*. New York, NY: Oxford University Press.
- Nagel, T. (1974). What Is it Like to Be a Bat?, *Philosophical Review* 1974, 435–50.
- Nagel, T. (1986). *The View from Nowhere*. New York, NY: Oxford University Press.
- Neisser, U (1993) *The Perceived self: Ecological and Interpersonal Sources of Self Knowledge* New York, NY: Cambridge University Press
- Newen, A. and Schlicht, T. (2009). Understanding Other Minds: A Criticism of Goldman's Simulation Theory and an Outline of the Person Model Theory. *Grazer Philosophische Studien* 79 (1): 209–242.
- Noe, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.
- Noe, A. (2009) *Out of Our Heads*. NY, NY: Hill and Wang
- Noonan, H. W. (1981). Methodological Solipsism *Philosophical Studies* 40: 269–274.
- Nozick, R. (1981). *Philosophical Explanations*. Cambridge, MA: Harvard
- Nuccitelli, S. (2003). *New Essays on Semantic Externalism and Self-Knowledge*. Cambridge, MA: MIT Press.
- O'Brien, L. (2007). *Self-Knowing Agents*. New York, NY: Oxford University Press
- Peacocke, C. (1983). *Sense and Content*. New York, NY: Oxford University Press

- Peacocke, C. (2003). "Action: Awareness, Ownership and Knowledge" In *Agency and Self-Awareness*. Ed. Roessler and Eilan Oxford, UK: Oxford University Press.
- Perry, John (1979). The Problem of the Essential Indexical. *Self-Knowledge*. Ed. Cassam New York, NY: Oxford University Press
- Piaget, J. (1955). *The Child's Construction of Reality*. London, UK: Routledge and Kegan Paul.
- Piston, A. E. (2002). *Hume's Philosophy of the Self*. New York, NY: Routledge
- Putnam, H. 1975. The meaning of "meaning". In K. Gunderson (Ed.), *Language, Mind and Knowledge*, Vol. VII of *Minnesota Studies in the Philosophy of Science*, 131-193. University of Minnesota Press.
- Quine, W. V. O. (1953) Two Dogmas of Empiricism In *From a Logical Point of View* Cambridge, MA: Harvard University Press, pp. 20–46.
- Reddy, V. (2005). Before the Third Element. In Eilan, N., Hoerl, C., McCormack, and Roessler, J. (Eds). *Joint Attention: Communication and Other Minds*. (pp. 85–109). Oxford, UK: Oxford University Press.
- Robbins, R. (Forthcoming). Self-Knowledge: A Brain's Eye View.
- Rochat, P. (2001). *Others in Mind*. Cambridge, UK: Cambridge University Press.
- Rodl, Sebastian (2007). *Self-Consciousness*. Cambridge, MA: Harvard University Press.
- Roessler, J. (2002). Action, Emotion, and the Development of Self-Awareness. In Elisabeth Pacherie (Ed.). *European Review of Philosophy: Emotion and Action*. Vol. 5. Stanford, CA: CSLI Publications.

- Roessler, J. (2005). Joint Attention and Other Minds. In Eilan, N., Hoerl, C., McCormack, and Roessler, J. (Eds). *Joint Attention: Communication and Other Minds*. Oxford, UK: Oxford University Press.
- Rosenberg, J. (1975). Transcendental Arguments Revisited *Journal of Philosophy* 72: 611–24
- Rosenberg, Jay (1986). *The Thinking Subject*. Philadelphia, PA: Temple University Press
- Rosenthal, D. (2000). *Materialism and the Mind-Body Problem*. Indianapolis, IN: Hackett Publishing.
- Rosenthal, D. (2005). *Consciousness and Mind*. New York, NY: Oxford University Publishing.
- Rowlands, M. (1995). Against Methodological Solipsism: The Ecological Approach. *Philosophical Psychology* 8.1: 5–24.
- Russell, J. (1995). At Two with Nature: Agency and the Development of Self-World Dualism. In J. L. Bermúdez, A. Marcel and N. Eilan (Eds.), *The Body and the Self* (pp. 127–151). Cambridge, MA: MIT Press.
- Ryle, G. (1949). *The Concept of Mind*. Introduction by Daniel Dennett Chicago, IL: University of Chicago Press.
- Schwitzgebel, E. (2008). The Unreliability of Naive Introspection *Philosophical Review* 117: 245–273
- Searle, J. (1983). *Intentionality*. Cambridge, UK: Cambridge University Press.
- Searle, J. (2004). *Mind: A Brief Introduction*. New York, NY: Oxford University Press.
- Seemann, A. (2008). Person Perception *Philosophical Explorations* 11.3: 245–262.
- Seemann, A. (2010). *Joint Attention: New Developments*. Cambridge, MA: MIT Press.

- Sellars, W. (1956). *Empiricism and the Philosophy of Mind* Ed. Robert Brandom.
(1997). Cambridge, UK: Harvard University Press.
- Sellars, W. (1963). Philosophy and the Scientific Image of Man. In *Science, Perception and Reality*. New York, NY: Routledge and Kegan Paul. pp. 1–40.
- Sellars, W. (1969). Metaphysics and the Concept of a Person in *Essays in Philosophy and its History*. Dordrecht, Holland: D. Reidel Publishing Co 1975: 214–241
- Sellars, W. (1970). ... this I or he or it (the thing) which thinks the 1970 Presidential Address, American Philosophical Association (Eastern Division), in *Essays in Philosophy and its History*. (Dordrecht, Holland: D. Reidel Publishing Co) 1975: 62–90
- Shoemaker, Sydney (1968). Self-Reference and Self-Awareness. *Self-Knowledge*. Ed. Quassim Cassam New York, NY: Oxford University Press
- Shoemaker, Sydney (1994). Self-Knowledge and “Inner Sense.” *Philosophy and Phenomenological Research*, (54)2, 249–314
- Shoemaker, Sydney (1996). *The First Person Perspective and Other Essays*. New York, NY: Cambridge University Press
- Sluga, Hans (1996). Who’s house is that? Wittgenstein on the Self. *The Cambridge Companion to Wittgenstein*. Cambridge, UK: Cambridge University Press
- Smith, J. (2006). Which Immunity to Error? *Philosophical Studies*, 130, 273–283.
- Smith, J. (Diss) *Self-Consciousness and Embodied Experience*. Submitted 2003
Department of Philosophy, University College London.

- Stich, S. (1980). Paying the Price for Methodological Solipsism In *The Nature of Mind*.
Ed. David Rosenthal. New York, NY: Oxford University Publishing. pp. 499–
500
- Strawson, G. (2008). *The Grounds of Self-Consciousness*
- Strawson, G. (2009). *Selves: An Essay in Revisionary Metaphysics*. New York, NY:
Oxford University Press.
- Strawson, P. F. (1959). *Individuals*. New York, NY: Taylor and Francis.
- Strawson, P. F. (1966a). *The Bounds of Sense*. New York, NY: Routledge.
- Strawson, P. F. (1966b). Self, Mind and Body. *Common Factor*. 4: 5–13.
- Strawson, P. F. (1985). *Skepticism and Naturalism*. London, UK: Methuen & Co.
- Strawson, P. F. (1994) Reference and the First Person. In Cassam, Q. (Ed.) *Self-
Knowledge*.
- Stroud, B. (1968) Transcendental Arguments *Journal of Philosophy* 65
- Stroud, B. (1999) The Goal of Transcendental Arguments in Robert Stern (Ed.).
Transcendental Arguments: Problems and Prospects. Oxford, UK: Oxford
University Press.
- Taylor, C. (1975). *Hegel*. Cambridge, UK: Cambridge University Press.
- Thompson, E. (2001). *Between Ourselves*. Thorverton, UK: Imprint Academic.
- Tomasello, M. (1993). On the interpersonal origins of the self-concept. In Neisser, U.
(Ed.) *The Perceived Self*. Cambridge: Cambridge University Press.
- Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. Dunham
(Eds.). *Joint attention: Its origins and role in development* (pp. 103–130).
Hillsdale, NJ: Lawrence Erlbaum.

- Tomasello, M., Call, J. & Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Animal Behaviour*, 55, 1063–1069.
- Trevarthen, C. (1979). Communication and Cooperation in early infancy: a description of primary intersubjectivity in M. Bullowa (Ed.) *Before Speech*, pp. 336–340.
- Trevarthen, C. and P. Hubley. (1978). Secondary Intersubjectivity: Confidence, Confiding, and Acts of Meaning in the First Year. In Lock, A. (Ed.) *Action, Gesture, and Symbol: The Emergence of Language* (pp. 183–229). New York, NY: Academic Press.
- Tugendhat, E. (1986). *Self-Consciousness and Self-Determination*. Trans. Paul Stern
Cambridge, MA: MIT Press
- Wiggins, D. (1980). *Sameness and Substance*. Cambridge, MA: Harvard University Press.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell
- Wittgenstein, L. (1958). *The Blue and Brown Books*. New York, NY: Basil Blackwell
- Wittgenstein, L. (1961). *Tractatus Logico-Philosophicus*. New York, NY: Routledge and Kegan Paul
- Wright, C. (1998). Self-Knowledge: The Wittgensteinian Legacy In *Knowing Our Own Minds*. Ed. Cripsin Wright, Barry C. Smith, and Cynthia Macdonald. New York, NY: Oxford University Press, pp. 13–46.
- Xu, F. (2007). Sortal Concepts, Object Individuation and Language. *Trends In Cognitive Sciences*. Vol. 11.9, 400–406.
- Zahavi, D. (2007). *Subjectivity and Selfhood*. Cambridge, MA: MIT Press.