

2016-08-16

Making Ourselves Intelligible - A Nondescriptivist Approach to Propositional Attitudes

Sarah B. Lesson

University of Miami, sarahlesson@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation

Lesson, Sarah B., "Making Ourselves Intelligible - A Nondescriptivist Approach to Propositional Attitudes" (2016). *Open Access Dissertations*. 1728.

https://scholarlyrepository.miami.edu/oa_dissertations/1728

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.

UNIVERSITY OF MIAMI

MAKING OURSELVES INTELLIGIBLE – A NONDESCRIPTIVIST
APPROACH TO PROPOSITIONAL ATTITUDES

By

Sarah Beth Lesson

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

August 2016

©2016
Sarah Beth Lesson
All Rights Reserved

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

MAKING OURSELVES INTELLIGIBLE – A NONDESCRIPTIVIST APPROACH
TO PROPOSITIONAL ATTITUDES

Sarah Beth Lesson

Approved:

Amie Thomasson, Ph.D.
Associate Professor of Philosophy

Mark Rowlands Ph.D.
Associate Professor of Philosophy

Eli Chudnoff, Ph.D.
Assistant Professor of Philosophy

Guillermo J. Prado, Ph.D.
Dean of the Graduate School

Julia Tanney, Ph.D.
Professor of Philosophy

Kristin Andrews, Ph.D.
Associate Professor of Philosophy
York University

LESSON, SARAH BETH
Making Ourselves Intelligible – A Nondescriptivist
Approach to Propositional Attitudes

(Ph.D., Philosophy)
(August 2016)

Abstract of a dissertation at the University of Miami

Dissertation supervised by Professor Amie Thomasson
No. of pages in text. (166)

In this dissertation, I argue for a function-first account of propositional attitude discourse that challenges a predominant assumption in philosophy of mind, that propositional attitude terms aim to refer – whether to brain states, functional states, dispositions, or entities posited to explain and predict. Using empirical evidence to support theoretical argumentation, I develop the Intelligibility View of propositional attitude discourse that highlights the distinctively *normative* role that the attribution of such states plays in social interaction. Building off the Intelligibility View, I argue that some of the most entrenched problems in this domain, including the problem of mental causation, problem of other minds, and the placement problem, all arise by accepting the descriptivist assumption and can be avoided by rejecting descriptivism. I conclude by showing how we can remain wholly realist about propositional attitudes once we understand the proper function of the relevant discourse.

For Nana, without whom nothing would be possible

Acknowledgments

I'd like to thank my committee members for their patience, encouragement, and support – Eli Chudnoff, Mark Rowlands, Kristin Andrews, Julia Tanney, and especially Amie Thomasson, whose writings and lectures fanned the flames of my philosophical minimalism, and inspired me to think about the mind in a wonderfully unpopular way.

Thanks to my family, especially my sister, Jenny Thomas. I've always aspired to be the person she thinks I am, making her proud is the most meaningful thing I have achieved.

I want to thank my wonderful friends – Heleana Theixos, Barry Devine, Sabrina Bano Jamil, and Ryan Lake for not giving up on me even at my most obnoxious, being there when I needed them, and more importantly being lovingly not-there when I needed to hermit away.

I'd also like to thank Uriah Kriegel, for his time, guidance, and kindness. And finally, thanks and love to Steven Olson. Words can't express how lucky I am that we met each other when we did. I'd never have finished this dissertation without the security and stability you bring to my life, deer.

Table of Contents

CHAPTER 1: INTRODUCTION
1

CHAPTER 2: HISTORICAL ROOTS
11

CHAPTER 3: NONDESCRIPTIVISM AND THE PROPOSITIONAL ATTITUDES
48

CHAPTER 4: THE EMBEDDING PROBLEM AND OTHER OBJECTIONS
88

CHAPTER 5: EASY ONTOLOGY AND THE MIND
118

CHAPTER 6: CONCLUSION
148

BIBLIOGRAPHY
161

CHAPTER 1: INTRODUCTION

The central aim of this project is to develop a nondescriptivist approach to propositional attitude talk and use this account of the language we use to discuss beliefs, desires, doubts, hopes, and intentions as a starting place to defend a simple realism about those state's ontology. By a *nondescriptivist* approach to propositional language talk, I mean a function-first exposition of mental discourse which claims that the foundational purpose of the relevant terms is not to describe, but instead to play the nondescriptive, normative, function of making human behavior intelligible. However, ultimately the view I advocate for *does* permit a deflated sense in which we come to use propositional attitude terms to describe mental states, so let me clarify the sense of *descriptive* I reject. In his 2008 *Two Readings of Representationalism*, Huw Price makes a helpful distinction between what he calls *e-representational* (external) states and *i-representational* (internal) states¹. E-representations are “environment-tracking” relations, dependent on a notion of covariation. The bar in the thermometer raises when the temperature does the same; a tree gains another growth ring roughly each year as changes in moisture and climate reliably affect new cells. With e-representations there is some intended or actual parallel variation between what is being tracked and what is tracking. On the other hand, there are i-representations, where something counts as a representation based on its role in a cognitive or inferential system. According to Price, e-representations are characterized by a causal tracking relation, while i-

¹ Price, H. (2008). *Two Readings of Representationalism*. *Descartes Lecture given at Tilburg August, 5*. Also discussed in (2011). *Naturalism without mirrors*. Oxford University Press.

representations are characterized by a logical inferential relation. When I say that propositional attitude discourse is nondescriptive, I am denying that terms like “belief” “desire” and “intention” are e-representations, functioning to reliably track and correspond to a certain kind of entity. There is room, though, to say that such terms are i-representational, in the sense that they can meaningfully feature in declarative statements by deriving their meaning from their place within our linguistic and conceptual practices.

I will argue in chapter 2 that there exists both theoretical and empirical evidence that the practice of appealing to our own minds, particularly propositional attitudes, affords unique social abilities related to fostering cooperation, facilitating harmony by ameliorating group tension, expediting technological advancement, and underpinning more sophisticated normative practices of morality. This is not an entirely novel approach to understanding propositional attitudes; I draw heavily from the work of Julia Tanney and Kristin Andrews, whose own research draws on existing approaches in analytic philosophy and cognitive science. My positive contribution is to unify these, and other historical and contemporary projects, under a single heading of “nondescriptivism” and then to draw out the epistemological and metaphysical entailments of that meta-linguistic account. I call this claim about the normative nature of propositional attitude attribution the *Intelligibility View* – referring to propositional attitudes is a way of making human behavior intelligible in ways that advance group living. From the starting point of language, I then move to examine what we can say about the existence and nature of propositional attitudes given the Intelligibility View. After addressing some objections to the

semantic element of my project in Chapter 4, I move on to the ontology of propositional attitudes in Chapter 5, defending a *simple realism* towards these states. Simple Realism is an ontological strategy developed by Amie Thomasson², aiming to provide easy answers to ontological questions by appealing to competence with the relevant concepts. I will show how nondescriptivism about propositional attitude terms lends itself naturally to a simple realism towards the ontology of paradigmatic propositional attitudes like belief, desire, and intention. Taking as our starting point questions of how propositional attitude language *functions*, I argue, provides a promising path forward into questions about the existence and epistemology of these philosophically problematic entities. Approaching epistemic and ontological questions from the direction of language can shed light on hidden assumptions and issues that might otherwise go unseen. Taken together, the meta-linguistic nondescriptivism and metaphysical simple realism enable us to look at longstanding problems, like that of self-knowledge or mental causation, with a different emphasis. Now we're better able to notice presuppositions about how we use propositional attitude terms in our everyday lives, and how these presuppositions may be giving rise to problems downstream. For instance, one of the chief problems that arise from assuming that the language of propositional attitudes primarily functions to *track* a certain kind of thing in the world, is that worldly things have causal connections with other worldly things –

² Thomasson, A. L. (2015). *Ontology made easy*. Oxford University Press, and (2010). *Ordinary objects*. Oxford University

and so the problem of mental causation is necessitated by this assumption regarding the function of the language.

Motivating my project is the idea that approaching topics in philosophy of mind from the direction of language causes recurring debates like the mind/body problem, problem of other minds (and related puzzles about self-knowledge), and problems of mental causation/interaction to show up in a different light. The nondescriptivism towards propositional attitude terms that I argue for arises from an investigation of the sociolinguistic and conceptual practices related to propositional attitude ascription. In Chapter 2, I draw upon the wealth of recent research on reason explanation and false belief tests in order to demonstrate that propositional attitude ascription is not necessary, nor even typical, in everyday explanations of behavior. This leads me to question what the *distinctive* function of propositional attitude ascription might be; that is, what uniquely does having propositional attitude discourse at our disposal *do* for us? Here I elaborate upon Kristin Andrews' critique of cognitive science's emphasis on the explanation/prediction function of folk psychology, which she delivers a sustained argument against in *Do Apes Read Minds*³. Andrews presents compelling reasons to accept that propositional attitude ascription is unnecessary for explaining and predicting behavior, the central claim of the predominant view of folk psychology in contemporary cognitive science. Drawing upon studies from social, developmental, and cognitive psychology Andrews urges that,

³ Andrews, K. (2012). *Do apes read minds?: Toward a new folk psychology*. MIT Press.

...We think about beliefs in particular kinds of situations, such as when a person deviates from expected behavior or violates the norms of society, but we don't need to appeal to beliefs to predict quotidian behavior. (2012, p. 10)

This passage illustrates a theme important in Andrews work, and my own. Insofar as any theory attempts to incorporate a commonsense understanding of the mind, it's imperative that it gets commonsense psychology *right*. Regardless of whether 'the folk' are right or wrong in their practical conception of the mind, if a view in philosophy or psychology takes folk psychology as its starting point, it needs to be careful in representing folk concepts how they're actually used. Andrews extensively demonstrates ways in which folk psychology has been misunderstood by theorists, and I draw upon her arguments in Chapter 2 to argue that the failure to acknowledge the functional pluralism of propositional attitude discourse directly leads to the metaphysical placement and interaction problems, and epistemological worries about the mechanism of self-knowledge. Adding to the empirical concerns about an overly simplistic descriptivist assessment of propositional attitude ascription, I also offer some theoretical shortcomings of the view. Here I confront Donald Davidson's influential argument from *Actions, Reasons, and Causes*⁴ for why reason explanations must be understood causally. While there are a number of considerations that lead to the conclusion, chief among them is the insight that,

How about the other claim: that justifying is a kind of explaining, so that the ordinary notion of cause need not be brought in? Here it is necessary to decide what is being included under justification. Perhaps it means only...that the agent has certain beliefs and attitudes in the light of which the action is reasonable. But then something essential has certainly been left out, for a person can have

⁴ Davidson, D. (1963). Actions, reasons, and causes. *The journal of philosophy*, 60(23), 685-700.

a reason for an action, and perform the action, and yet this reason not be the reason why he did it. Central to the relation between a reason and an action it explains is the idea that the agent performed the action because he had the reason. (1963, p. 691)

Julia Tanney has written at length detailing weaknesses in Davidson's handling of this issue⁵. I invoke Tanney's compelling criticism which claims that the metaphysics of mind entailed by Davidson's argument are untenable, and therefore something has gone wrong with his argument. Specifically addressing the worry raised by Davidson in the preceding passage, Tanney says⁶,

We can concede this. But this does not show, as Davidson argues, that a causal relation is required between reason and action in order to secure explanation...it merely shows that what was one candidate for a sense-making patten should be rejected for another. (2013, p. 8)

Extending this point from the question of reason-explanation to the ways propositional attitude ascriptions factor in our explanations of behavior, I argue in Chapter 5 that a causal constraint on propositional attitudes is unnecessary for explaining how such terms are used in practice, and also unlikely to obtain given interesting findings in recent neuroscience. Given the combined facts that propositional attitudes are largely unnecessary for explanation/prediction of behavior and empirical evidence challenges the causally efficacy of such states, accepting Davidson's argument, indicative of the dominant view that the language of propositional attitudes must aim to refer to content-bearing causally efficacious

⁵ See Tanney, J. (1995). Why reasons may not be causes. *Mind & Language*, 10(1-2), 105-128, (2005). Reason-Explanation and the Contents of the Mind. *Ratio*, 18(3), 338-351

⁶ Tanney, J. (2013). *Rules, reason, and self-knowledge*. Harvard University Press.

states, might lead us to eliminativism about propositional attitudes. A key upshot of my semantic and metaphysical arguments is that we can reasonably reject the a causal constraint on the existence of mental states and avoid the peril of eliminativism.

Chapter 2 serves to identify a strain of nondescriptivism in philosophy of mind emerging in the first half of the twentieth century. I discuss three notable predecessors that demonstrate the methodological motivations for going nondescriptive, and assess what these early attempts got *right* and *wrong* when it comes to a function-first strategy of understanding the mind. I will consider specific writings of Wittgenstein, Ryle, and Sellars, thinkers who arguably offered the first attempts to study mental phenomena (from perception to propositional attitudes, and even consciousness) by taking a careful look at the relevant conceptual and linguistic practices. These early attempts sought to clarify the distinctive capacities such discourse affords us, and trace out the logical geography of mental predicates. This strategy of approaching issues in philosophy of mind through language was largely rejected during the rise of functionalism in the latter half of the twentieth century, for what I will argue were insufficient reasons. Because this work was wrongly interpreted through a lens of descriptivism, the respective accounts faced objections that led to them being judged as insufficient to bear the necessary epistemic and metaphysical burdens. We learn about these views now, if at all, as historically provocative, yet philosophically futile theories of mind. But by dismantling the objections that caused certain areas of work by Wittgenstein, Ryle, and Sellars to go neglected, and excavating the tacit assumption that some

type of descriptivism was lurking beneath the surface of their accounts, I hope to pave the way towards preserving and putting to use some of the more promising insights these early nondescriptivists supplied. Part of the significance of the propositional attitude nondescriptivism I propose here is that it helps make clear what went wrong in our dismissal of these historical accounts. One example of this is Ryle's theory advanced in *Concept of Mind* wrongfully being classified as the thoroughly refuted theory of logical behaviorism, a mistake that Julia Tanney has recently quite compellingly⁷ attempted to correct. The upshot of righting these historical wrongs is that we remove a barrier to embracing a promising strategy of explaining mental phenomena, the baby that had been tossed out with the bathwater of logical behaviorism is still alive and kicking. I will pick up where these views left off, building a conceptually rigorous account of propositional attitudes that elucidates confusions in contemporary debates. But there were also legitimate objections to this line of work—most notably the Frege-Geach problem. Bringing contemporary neo-pragmatist work (Price, Williams) to bear on the embedding problem first posed by Frege and Geach, I will argue that a functionally driven account of mental discourse is finally able to overcome what earlier, and simpler, versions of the view could not. Here I am indebted to contemporary nondescriptivist strategies in other areas of philosophy, such as Simon Blackburn's quasirealism of the moral, Amie Thomasson's take on modality, and Huw Price's global expressivism. By employing these recent advances, I argue that my function-first account of mental discourse leads naturally to an epistemically and

⁷ Tanney, J. (2009). *Rethinking Ryle: A Critical Discussion of The Concept of Mind*.

metaphysically sufficient account of propositional attitudes that clears up the misguided objections to earlier forms of nondescriptivism, while avoiding the substantive objections faced by those views.

This propositional attitude nondescriptivism aims to resolve some of the more puzzling problems in philosophy of mind, particularly the mind/body problem and the problem of mental causation, by turning a critical eye to how these questions are framed in the first place, arguing that the difficulty lies in the questions posed and not the phenomenon we're after at all. To demonstrate this point it will be necessary to analyze the descriptivist assumption about mental discourse, both in common language and philosophy. The strategy I will adopt is to first look to the function of propositional attitude terms to articulate their distinctive function, what unique moves such terms allow us to make in our linguistic practices. Particularly, I will argue that these terms function in a normative, justificatory, way to facilitate social harmony and ameliorate tensions that arise between people. This understanding of the function of propositional attitude talk, offers a promising route to understanding mental states through a study of the relevant conceptual and linguistic practices. The suggestion is that this approach is able to preserve some of the insights of prior non-descriptivist views while avoiding their problems. In closing, I begin by examining the consequences of adopting the Intelligibility View and subsequent simple realism. Though the view is most decidedly *not* a type of eliminativism, dropping the descriptivist assumption leads to a provocative view of how talk about our beliefs, desires, and intentions functions, what we can say about the nature of such mental states, and of which

sorts of problems are genuine, as opposed to being byproducts of looking at the discourse with the wrong assumptions.

Ultimately, the proposed project has both historical and contemporary import. By defending an often-overlooked strategy from twentieth century analytic philosophy against misguided criticism, I will formulate a meta-ontological theory of propositional attitudes that ambitiously aims to resolve (by explaining away) what seemed like intractable problems in contemporary philosophy of mind.

CHAPTER 2: HISTORICAL ROOTS

Overview

In recent years, a philosophical strategy popularized throughout the early and mid twentieth century has received renewed interest among philosophers. The general idea is that by taking a language-centered route into philosophical questions, the sociolinguistic/conceptual practices of a certain domain can provide an advantageous entry point for understanding metaphysical or epistemological problems in a new way. More specifically, the strategy is to approach classic puzzles involving some class of disputed entities (modal or moral facts, numbers, propositions, properties, etc.) and question a widespread assumption about the way the discourse involving terms for those entities *functions*. The assumption is that the language involving the disputed entities aims to *describe* some feature or entity in the world, and the nondescriptivist challenges that assumption by carefully tracing out what the relevant conceptual/linguistic practices allow us to *do*. Language plays many roles for us other than to describe, for example, we can warn, admonish, express, console, and obfuscate. If the nondescriptivist can provide a compelling argument for why we should think some domain of discourse doesn't primarily aim to describe, important epistemic and metaphysical results might surface. We find early examples of this strategy in Ayer's ethical expressivism (1954) and deflationism about truth (1936), Wittgenstein's expressivism about sensation-talk (1958), Sellars speech act treatment of appearances and account of physical laws as inference tickets (1963), and Ryle on mental states and conditionals (1949). If the examination of the relevant domain

of language turns up a foundationally *nondescriptive* function, the strategy then leads into an investigation of what the misunderstanding of the function of the discourse might mean for philosophical issues downstream.

While nondescriptivism has enjoyed a recent resurgence, the popularity of the approach has not yet taken root in philosophy of mind. There are probably a number of reasons why this is the case. First, some areas of discourse just intuitively seem like better candidates for a nondescriptive treatment than others. Take property-talk, for instance. While of course there are age-old and interesting debates about the existence and nature of properties, on initial inspection property-talk may be thought to be more amenable to a nondescriptivist treatment in a way that talk about things like beliefs, conscious awareness, and pains is not. Property-talk simply isn't as indispensable to our everyday lives as talk about what we believe and want. Stephen Schiffer advances what he calls a *pleonastic* account of both properties and propositions in his 2003 *The Things We Mean*⁸. A similar strategy can be found in Crispin Wright (1983) and Bob Hale's (with Crispin Wright 2001, 2009) neo-Fregean account of numbers and Amie Thomasson's view on ordinary objects (2007). When it comes to mental discourse there seems to be less intuitive force behind the claim that terms such as 'belief,' 'desire,' and 'intention' don't aim to describe things we encounter in contact with, either directly or indirectly. A contributing factor in the reluctance towards nondescriptivism in philosophy of mind might also be its association, wrongly or rightly, with eliminativism. I will

⁸ Schiffer, S. (2003). *Things We Mean*. Oxford University Press

confront these worries in later chapters, ultimately rejecting either as a serious obstacle for the type of nondescriptivism I develop here.

Strands of mentalistic nondescriptivism were developed by Ryle, Sellars, and Wittgenstein in mid twentieth century, and enjoyed a flash of popularity before being unilaterally rejected due to what were taken to be insurmountable problems raised by a host of critics in the 1960s. The most notable problem, articulated by Gottlob Frege, refined by Peter Geach (1965), and arrived at independently by John Searle (1962) is the *embedding problem* (often called the Frege-Geach problem), which establishes that the semantic content of a given term cannot be equated with its assertoric force. This objection originally was put to non-cognitivist approaches to morality that grew popular in the first half of the twentieth century [Ogden & Richards (1923), Barnes (1933), Stevenson (1937, 1944), Hare (1952), and Ayer (1952)]. In different ways, these views advanced the idea that moral judgments do not *function* like typical declarative statements, and instead of aiming to reference properties in the world like *moral goodness* or *badness*, the terms were expressive in nature. Mark Schroeder distills the main force of the embedding problem in his 2008 article *What is the Frege-Geach Problem*⁹, saying

...Both Geach and Searle were pressing the objection that non-cognitivists are committed to denying that 'good' or 'wrong' mean the same thing in at least certain kinds of embedded contexts as they do in simple atomic sentences, and that this is bad, because we *need* to assume that they mean the same thing in both places, in order to explain the semantic properties of the complex sentences. (2008, p.706)

While initially leveled at moral non-cognitivism, the force of the embedding problem

⁹ Schroeder, M. (2008). What is the Frege-Geach Problem? *Philosophy Compass*, 3(4), 703-720.

spread from ethics into other areas of philosophy where nondescriptivist projects had emerged, and in part led to the rejection of those accounts and the strategy in general. In this chapter, I will highlight the nondescriptivist strains in Ryle, Sellars, and Wittgenstein, discussing the motivations behind their respective projects, and the promising insights of each. This will put me in a good position to explore the objections that lead to early versions of nondescriptivism being widely rejected, centrally how the embedding problem arises for each. This discussion will serve as the starting point for a new and thoroughgoing nondescriptivism about propositional attitudes able to overcome existing objections while preserving important historical insights developed in subsequent chapters.

§2.1 Ryle and the Mind

Although the categorization of *descriptivist/nondescriptivist* isn't explicitly used within the work, the strain of thinking developed in Ryle's *Concept of Mind*, and work advanced in subsequent essays, is inarguably one of the earliest views to challenge descriptivist assumptions in philosophy of mind. In what follows, I will draw out this element of his work, focusing on Ryle's famous insight that taking the mind to be a special sort of *thing*, appropriate for comparison and contrast with *physical things*, is a pervasive category mistake contaminating many popular debates within philosophy of mind. I will draw out Ryle's insistence that focusing on the language we use to talk about our minds is the key to unlocking corresponding epistemological and metaphysical puzzles. In his writings about the

mind, Ryle showed that despite the fact that nearly no one defends Cartesian Dualism anymore, substantive Cartesian assumptions still have a dominant role in shaping the questions we ask about the mind and the resulting answers we're able to give to those questions. Sadly enough, what was true in 1949 at the time when *Concept of Mind* was published is still just as true today, six decades later. While we can agree that few philosophers defend the now simplistic view Ryle disparagingly called the *official doctrine* of the 'ghost in the machine,' it is important to step back and see how elements of that view remain pervasive throughout contemporary debates in philosophy of mind. In the foreword to the 60th Anniversary Edition of *Concept of Mind*, Julia Tanney distills from Ryle's work three distinctive commitments that a broadly *Cartesian* approach entails, and how these commitments still muddle our thinking about the nature of the mind. Tanney tells us that the three commitments (ontological, epistemological, and semantic) are as follows:

- **Ontological commitment:** the ontological strand of the view is that there are two different kinds of things, body and mind, that are somehow harnessed together. The one exists in space and is subject to mechanical or physical laws and the other one is not in space and is not subject to these laws. And yet, the mind and body influence each other.
- **Epistemological commitment:** Mental processes or events are supposed, on the official view, to be played out in a private theatre; such events are known directly by the person who has them either through the faculty of introspection or the 'phosphorescence' of consciousness...Others can know them only indirectly through 'complex and frail inferences' from what the body does.
- **Semantic commitment:** mental terms function to name

phenomena that the epistemological aspects of the doctrine assure us are hidden¹⁰

These commitments extracted from Ryle's work and articulated by Tanney, are essential to understanding a central motivation for going nondescriptive about mental discourse. The standard story, familiar to us far before we ever take our first philosophy class, is that the mind is a thing stored (figuratively or literally) inside of us), itself containing smaller things that we call 'beliefs' 'desires' 'emotions' and 'memories.' Starting from this metaphysical understanding, we're prompted to ask questions about how we can know these things – we clearly don't see them with our eyes or *hear* them with our ears. We then set off constructing elaborate theories about the fundamental nature of such entities, and how they come to be known, often introducing further entities – like the faculty of introspection, sense data, and phenomenal properties - in the service of our original explanations, which result in further metaphysical and epistemological questions. All of this hinges on the underlying semantic commitment – that the language, in the first place, really functions to describe this class of entities and processes. As Tanney says,

The underlying semantics of the official doctrine takes the meaning of mental expressions to be determined in part by what the words in such expressions name or designate in a similar way as the meaning of the sentence 'Jones bought the most expensive house in the village' is given in part by the individual named by 'Jones' and by the particular piece of real estate picked out by the definite description 'the most expensive house in the village.' (2009, p. xvii)

¹⁰ Tanney, J. (2009). *Rethinking Ryle a critical discussion of The Concept of Mind*. Routledge, p. xi-xvii

But this descriptivist assumption gives rise to a bevy of philosophical puzzles, as Tanney explains,

Thus the semantic accoutrements of the official doctrine—the view that mental terms function to name phenomena that the epistemological aspects of the doctrine assure us are hidden—lead directly to the philosophical conundrum known as the threat of (necessarily) private languages for mental phenomena. (2009, p. xvii)

and the private language worry leads us directly into the problem of other minds – a problem that still sees no satisfactory answer in current debates,

The problem of other minds was at centre stage of discussions in philosophy of mind in the 1950s before the mind–body problem attracted the wider audience. The problem of other minds is this: if certain aspects of the official doctrine are correct and minds consist of episodes that are only privately knowable, then we need to rethink our claim to know (with certainty) that other minds exist. The thought at the time was that this was an intolerable conclusion, so philosophers set about to show how the claim to have knowledge of other minds is none the less justified. But though no longer at the centre, the problem of other minds lurks in the background of recent discussions of ‘phenomenal consciousness’, which inherit the epistemological and semantical aspects of the official doctrine. (2009, p. xviii)

Since a number of trouble problems find root in a descriptivist assumption about mental state terms, it is reasonable to wonder if rejecting that assumption might possibly help us avoid the worries altogether. Insofar as these commitments are still widely endorsed today, a nondescriptive approach to mental discourse continues to be a worthwhile strategy to explore.

In order to appreciate Ryle’s idiosyncratic views on the mind, it is best to begin where he did – with the goal of refuting the official doctrine. Ryle begins *Concept of Mind* by saying,

There is a doctrine about the nature and place of minds which is so prevalent among theorists and even among laymen that it deserves to be described as the official theory. Most philosophers, psychologists, and religious teachers subscribe, with minor reservations, to its main articles and, although they admit certain theoretical difficulties in it, they tend to assume that these can be overcome without serious modifications being made to the architecture of the theory. The official doctrine, which hails chiefly from Descartes, is something like this. With the doubtful exception of idiots and infants in arms every human being has both a body and a mind. His body and his mind are ordinarily harnessed together, but after the death of the body his mind may continue to exist and function...

A person is also generally supposed to be able to exercise from time to time a special kind of perception, namely inner perception, or introspection. He can take a (non-optical) 'look' at what is passing in his mind. Not only can he view and scrutinize a flower through his sense of sight and list to and discriminate the notes of a bell through his sense of hearing; he can also reflectively or introspectively watch, without any bodily organ of sense, the current episodes of his inner life...On the other side, one person has no direct access of any sort to the events of the inner life of another. He cannot do better than make problematic inferences from the observed behavior of the other person's body to the states of mind which, by analogy from his own conduct, he supposes to be signaled by that behaviour. Direct access to the workings of a mind is the privilege of that mind itself; in default of such privileged access, the workings of one mind are inevitably occult to everyone else.¹¹

We can see in this passage the three commitments discussed above at play in Ryle's description of the target. Minds and mental states are *things*, special things known in a special way. And if that's the case, then the language we use that involves these things must aim to *describe*, just as discourse involving flowers and ringing bells aims to describe. This connects back to Price's distinction between e-representations and i-representations; the standard story in philosophy of mind is that mentalistic terms are *e-representations* in the sense that they aim to track

¹¹ Ryle, G. (2009). *The concept of mind*. Routledge, p. 1-4

things in the environment with which they covary. And it is this commitment that the nondescriptivist denies; claiming instead that propositional attitude terms are better understood as *i-representations*, deriving their meaning not from correspondence to things in the world, but instead from the inferential relations they bear to other terms and concepts. The metaphysical and epistemological commitments are explicit in the official doctrine, while the semantic commitment is an assumption necessitated by the former. If beliefs, desires, and hopes are *things* in the world, and we come to know them through some special inner sense, then semantically we are committed to descriptivism – to an account of propositional attitude terms as *e-representations*.

The sort of view that Ryle targets in this extended passage is not simply Dualism. As Tanney points out,

the official doctrine is dead in only one of its ontological aspects: substance dualism may well have been repudiated but property dualism still claims a number of contemporary defenders. Indeed, both non-reductive and reductive physicalists are entangled in a metaphysical overgrowth whose roots are firmly established in the soil of the official doctrine. The problem of finding a place for the mental in the physical world, of accommodating the causal power of the mental, and of accounting for the phenomenal aspects of consciousness are all live problems in the philosophy of mind today because they share some combination of the doctrine's ontological, epistemological, and semantic assumptions. (2009, X)

The broadly Cartesian orientation picked out by Tanney still frames questions about the mind/mental states as *things* we possess and can come to know (often by gazing inwardly). This orientation to the subject matter of the mind still structures most of the major positions one can adopt towards mental phenomena – even those that attempt to criticize aspects of the paradigm still tacitly accept the

overall framing. These views include property dualism, reductive materialism, non-reductive materialism, functionalism, and even eliminativism. It is my contention that what is shared by these otherwise disparate positions is the underlying commitment to descriptivism. Once it's accepted that mental discourse functions to describe, we're constrained in the type of answers we can give to questions involving the mind. For this reason I will abandon talk of Cartesian assumptions, and discuss instead the *descriptivism* that lies at the foundation of the problems Ryle articulates. In my view, it's not that our metaphysical and epistemological commitments necessitate a descriptivist take on mental discourse. Just the opposite, it seems that the assumption of descriptivism toward mental discourse, particularly propositional attitude talk, misguidedly pushes us into confused metaphysical and epistemological views about the mind. When we take language involving apparent reference to beliefs, desires, and intentions to serve a tracking function analogous to perceptual object talk, the analogy structures our expectations about the types of things these mental states must be. We are well served to pull the assumption into the light of day to understand the role it plays in our talk about mental phenomena and the way we frame questions and debates centered on that topic. The fact that mental discourse aims to describe certain 'things' in the world is rarely argued for, but instead a wide range of views within philosophy of mind take this point to be so self-evident it need not even be mentioned. This point is illustrated by taking a look at the most discussed topics in philosophy of mind right now, such as mental causation, the explanatory gap

problem, the higher order vs. lower order thought debate, and computation vs. representation, among others.

In an extended attack on the official doctrine, Ryle argues that descriptivism, what he identifies as the assumption that language involving the mind, and its 'contents,' primarily has the purpose of *reporting*, has serious semantic, ontological, and epistemological consequences. Starting with the ontological claim of the official doctrine, there must exist a special class of *objects* that will serve in explanations of phenomena we are interested in. This assumption is descriptivist in the sense that it finds purchase in the comparison between discourse involving psychological phenomena on the one hand, and the language we use to report on the physical objects and events we perceive with our senses. This assumption – that there are special kinds of mental *things* that will explain mentality to us – is dualistic in the sense that the entities are distinct from physical ones. While they presumably exist in time, they do not exist in space. The ontological claim leads to the classic Mind/Body problem, and in contemporary debates rears its head in debates about mental causation and discussions of the possibility of naturalizing the mind. Because when we model talk of the mental on talk of the physical we are left with the question of exactly *how* mental objects are like (and unlike) physical ones, and how to trace out the causal relations among these things.

Stemming from the assumption that mental discourse serves to pick out a special class of entities is an epistemological assumption about how we may come to believe in, or know, such things. One's own mental states and experiences,

according to the Official Doctrine, are essentially privately observable from one's first person perspective, and not publically observable in the way physical objects are. There is no objective, third-person perspective on another person's mental states in the way we can have third-person access to another person's foot. The epistemological assumption is that there is a deep asymmetry between how we know our own minds and how we know the minds of others, rooted in the metaphysical difference between mental things and physical things. And that in the case of our own minds, there is an important privilege and priority to the access we enjoy. This assumption gave birth to the Problem of Other Minds classically, and more contemporarily arguments that rely on the so-called explanatory gap between the physical and phenomenal – such as the Knowledge Argument and Zombie problem.

In order to carry the metaphysical and epistemological commitments outlined above, a certain semantic assumption is implicit within the official doctrine. According to Ryle, the ordinary ways people use mental discourse are logically inconsistent with the way the world would have to be if descriptivism were true. In order to carry and be consistent with the ontological and epistemological commitments tacit in the Official Doctrine, mental discourse must serve to describe a class of private, internal entities. But this assumption has incoherent consequences, according to Ryle. If 'belief' necessarily picks out some private and 'internal' state of a subject, then when we never could have gotten the linguistic practices associated with such things off the ground. The way that our language involving mental predicates works allows us to attribute beliefs, desires, and

intentions not only to ourselves, but also to other people. But if such terms denote states or processes belonging only to the other people's inner life, then we would not, in principle, be justified to make such attributions. Furthermore, only the subject of the relevant mental event would be able to let us know if the term was correctly or incorrectly attributed. But this impossibility of genuine and justified third-person attribution is undermined by our ordinary practices of mental state attribution, disagreement about the correctness of that attribution, and affirming or denying when these attributions have been made of us. Therefore, the descriptivist semantic assumptions must be rejected, and with them the associated ontological and epistemic assumptions that rely on them. The so-called 'truth-makers' of our mental-talk cannot be essentially private, 'internal' states or processes because it would make the quite ordinary way we talk about beliefs, desires, intentions, emotions and mysteriously impossible. In the foreword to the 60th anniversary edition of *Concept of Mind*, Julia Tanney tells us,

Ryle's criticism of the official doctrine begins by pointing out an absurdity in its semantic consequences. If mental conduct verbs pick out 'occult' causes then we would not be able to apply those verbs as we do, so something must be wrong with a theory of mental phenomena that renders so inadequate our everyday use of these verbs¹².

If 'belief' necessarily picks out some private and 'internal' state of a subject, then we never could have gotten the linguistic practices associated with such things off the ground. The way that our language involving mental predicates works allows us to attribute mental states such as knowledge, belief, desire, depression, and

¹² ¹² Ryle, G. (2009). *The concept of mind*. Routledge, p. xix

pain to other people. But if such terms denote states or processes belonging only to the other people's inner life, then we would have no way, in principle, to make such attributions. Further more, only the subject of the relevant mental event would be able to let us know if the term were correctly or incorrectly attributed. But the impossibility of genuine third-person attribution is undermined by our ordinary practices of mental state attribution, disagreement about the correctness of that attribution, and affirming or denying when these attributions have been made of us. Therefore, the descriptivist semantic assumptions must be rejected, and with them the associated ontological and epistemic assumptions. The so-called 'truth-makers' of our mental-talk cannot be essentially private, 'internal' states or processes because it would make the quite ordinary way we talk about beliefs, desires, intentions, and emotions mysterious (if not impossible). From this insight, a strategy to better understand both minds and discourse *about* minds suggests itself – we begin by looking at how the relevant domain of discourse functions, not just for individual speakers but within the greater language – what does the ability to attribute propositional attitudes allow us to *do*, that could not be done without such attribution? Nondescriptivism flips the official doctrine on its head. The official doctrine begins by presuming that mental states are private things analogous but markedly different than physical states that we can only know with certainty in the first person case, then postulating an inner sense by which we can come to know these states, all the while assuming that mental discourse functions to describe these things and this knowledge. A nondescriptivist analysis of propositional attitudes instead begins by taking note of how mental predicates function, what

distinctive role this language plays in our social practices. If the foundational function of the discourse is not a fact-stating function, which I will argue is the case with propositional attitude talk in chapter 2, then we ask what would have to be true epistemically and metaphysically to make sense of the functional account of the discourse. Before developing this view more fully, I will finish the historical trajectory of the strategy by highlighting a few other salient Rylean insights and moving on to Sellars' contribution to non-descriptive theories of mind.

One of the more interesting, and I think helpful, aspects of Ryle's approach to philosophy in general, one shared by Wittgenstein, is to see the proper role of philosophy as uncovering and examining places of conflict and confusion in our conceptual and linguistic practices. "We have now to operate *upon* what we ordinarily operate readily and unquestioningly *with*." Making this move – from working *with* words and concepts in our first-order language, to working *on* those words and concepts in a second-order language gives rise to new questions and confusions that were not previously apparent. This effect can be noted in all areas of philosophical inquiry. It often takes the form of positing new kinds of *objects* to serve in the explanations of the phenomena we're after, analogous to the way that physical objects such as atoms, molecules, and cells play explanatory roles in our explanations of the physical phenomena we're investigating.

Platonic forms, propositions, intentional objects, sense data were recruited to appease our professional hankering to have a subject matter of our own. (1971b, vii¹³)

¹³ 1971b, *Collected Papers*, volume 2, London: Hutchinson.

But the idea that our expressions have meaning only insofar as they represent, or stand for, things in the world should be rejected, according to Ryle. Language plays many roles, only one of which is to reference objects of which we can become aware. This non-descriptivist approach was taken up by various philosophers throughout the twentieth century, most notably in the Moral Expressivism of Ayer, Stevenson, Hare, Blackburn, and Gibbard. Expressivism is the view that some statements that seem to refer to certain entities on the grammatical surface actually play the role of expressing mental states of the speaker. In the case of moral statements we can understand claims such as, "It is wrong to steal," not as appealing to some kind of fact or property, 'the wrongness of stealing,' but instead as expressing an attitude of the speaker, that she condemns or disapproves of stealing. But expressivism shouldn't be taken to be the only kind of nondescriptivist approach to a domain of discourse. There are many functions of language, and describing and expressing are but two of these.

Ryle is clearly impressed with the plurality of uses language can be recruited for, a theme he comes back to throughout *Concept of Mind* and subsequent writings. And in parsing the different roles that Ryle argued mental discourse plays we can come to understand why his view is mischaracterized when it is taken to be *behaviorist* in nature. *Analytic Behaviorism*, the account that is most often attributed to Ryle, is a theory about the meaning of mental terms and concepts. The theory claims that the meaning, or referent, of any mental-predicate is some behavioral disposition, or loosely connected collection of behavioral tendencies in which a subject might engage in a given situation. For example,

when you attribute a desire for ice cream to another person you are *not* referring to some private and internal state/process of the subject. While Ryle was a pluralist about mental state talk, arguing that its linguistic and conceptual tools afford us the ability to do a number of new things in our social interactions with others, one of the central distinctive abilities which mental state talk affords us is that of *licensing inferences*. As Ryle puts it:

Dispositional words like 'know', believe', 'aspire', 'clever', and 'humorous' are determinable dispositional words. They signify abilities, tendencies or pronenesses to do, not things of one unique kind, but things of lots of different kinds. Theorists...are apt to notice this point, but to assume that there must be corresponding acts of knowing or apprehending and states of believing; and the fact that one person can never find another person executing such wrongly postulated acts, or being in such states is apt to be accounted for by locating these acts and states inside the agent's secret grotto. (1949, 118-119)

While Ryle takes mental talk to be dispositional, it's not clear that he takes mental language's purpose to be *describing* dispositions, in fact it's pretty clear that he doesn't. Instead, Ryle focuses on how a dispositional analysis of mental vocabulary allows us to license inferences about past and future behavior. In this sense, Ryle's emphasis on dispositions to behave is not introducing a new candidate for the truth-maker of propositions involving mental predicates. Instead, he's calling attention to the different (nondescriptive) uses dispositional talk serves, not in describing entities, but in licensing inference and justifying behavior. The role of dispositions in Ryle's analysis is to draw our attention to the primacy of behavior in deploying, justifying, and evaluating mental state language. Since dispositions to behave play such an important role in Ryle's early attempt at a

nondescriptivist strategy, I'll say a few things about how we are meant to understand dispositions in this context.

According to Ryle, dispositions come in two types: single-track and higher-grade¹⁴. Single-track dispositions pick out precise reactions between physical objects that are governed by highly reliable physical laws that are essentially tied to things like their microstructure. Examples of this include sugar's disposition to dissolve in water, or a pane of glass' disposition to shatter when struck with a certain degree of force. With single-track dispositions, specifying initial conditions will fully determine the outcome of any interaction. Mental dispositions of people, according to Ryle, are not *indefinitely-heterogeneous*. Using the case of knowledge and belief for illustration, Ryle says,

Epistemologists, among others, often fall into the trap of expecting dispositions to have uniform exercises. For instance, when they recognise that the verbs 'know' and 'believe' are ordinarily used dispositionally, they assume that there must therefore exist one-pattern intellectual processes in which these cognitive dispositions are actualized...flouting the testimony of experience.

The more complicated dispositions intimated at with mental predicates do not demonstrate uniformity, they are 'higher-grade' in the sense that an interaction of a person with a set of circumstances could turn out a number of ways given that person's beliefs, desires, or intentions. My belief that the ice is thin is consistent with a number of related behaviors – my avoiding the ice altogether, or favoring the sides of the pond over the center, my warning other skaters to stay clear of certain spots, my eyeing the rescue equipment often – my demonstration of some

¹⁴ Ibid. p. 32-34

or all of these behaviors is adequate grounds to judge that I believe the ice is thin. And while Ryle does not go into detail about why we find this difference between single-track and higher-grade dispositions, nondescriptivism can provide an analysis of this difference. Mental dispositions, according to Ryle, play an essential role in evaluating attributions of mental predicates, knowing when a term like 'belief' as been correctly applied or not, whereas single-track dispositions appear to play a more elementary role of predicting and explaining interactions. Although Ryle didn't have recent advances in nondescriptivist strategies available to him, in Chapter 2 I will show how we can incorporate Ryle's insights about the relevance of behavioral dispositions when it comes to making sense of mental-discourse, by subsuming dispositions to behave as the material inferential clause of a three pronged strategy to articulate a function-first account of the meaning of propositional attitude terms.

Going back to Ryle's discussion of dispositions, we must be careful not to *identify* the disjunctive disposition with the given mental state, for instance claiming something like the following: the belief that the ice is thin *just is* a disposition to {avoid the ice altogether, stay to the sides of the pond while avoiding the center, warn other skaters of danger, eye rescue gear, etc.} Making this sort of identification puts us directly back into the tangles that arise from accepting the descriptivist assumption. Particularly when we misunderstand Ryle's *nondescriptive* analysis of dispositional talk itself. This would be just one more descriptivist solution to the question of mental states, answering the question "what are mental states" with the solution "they are disjunctive dispositions to behave."

The nuance of this position led many to misread Ryle as an Analytic Behaviorist. A more nuanced explication of Ryle's project is to understand the relevant behavioral profiles to be the standards by which we judge whether a person has a certain belief (desire, intention, emotion, etc.). The disjunctive dispositions are the *criteria* by which we are able to make sense of attributions of mental predicates, not the referents of mental predicates. Ryle is not advocating a dispositional analysis of propositional attitudes, or any other mental states. By emphasizing the importance of behaviors in analyzing mentalistic language, his focus is in pointing out how we use behaviors to *settle disputes* or *answer questions* involving mental states of others and ourselves. For instance, if we say Susan *doubts* that Tylenol reduces pain, then we are issuing a license to make certain inferences about her past, present, and future behavior: e.g., that if she had a headache, she would not take Tylenol, that she will not offer us Tylenol if we were to complain of a headache, that she may argue with us if we claim Tylenol is a good analgesic. Her doubt is not a *thing* she possesses that we are describing. In saying she doubts this claim, we are licensing specific inferences about her behavior, and providing criteria for judging whether or not the doubt that Tylenol reduces pain can rightly be attributed to Susan. This will go for attributions of other mental states such as desire, belief, intention, etc., all of which, on Ryle's view, aim not to describe special inner entities, but rather to license inferences about behavior. Ryle is commonly misunderstood on this point, with critics taking him to either think of our mental state talk as describing behaviors, or criticizing him for failing to say what the truthmakers of our dispositional talk are. For example, D.M. Armstrong says,

...We need to go on to consider the question of the truthmaker for these dispositional truths. What is there in the world in virtue of which these truths are truth? Ryle had no answer. Once we do raise the truthmaker question, then our view of the nature of mind will very likely be transformed and we will move in a quite un-Rylean direction. We will (very likely) identify a belief, say, with some inner *state* of the mind (materialist metaphysicians will identify it further with some state of the brain) that, in suitable circumstances, but on in suitable circumstances, will manifest itself in various ways, some of which may be outward behavior. (2004, p. 2-3)

But the nondescriptivism that underlies Ryle's project means that Armstrong is really missing the point here, Ryle's point is that talk of mental states should not be understood indicatives that require truthmakers¹⁵. Moving forward, as I try to preserve the insight that understanding dispositional talk is the key to our understanding the logic of mental discourse, I will be careful not to take dispositions to be the *meanings* or *referents* of mental terms, otherwise we land squarely back in the tangle that descriptivism creates.

Ryle's brand of nondescriptivism has the benefit of avoiding the more serious objections brought against descriptivist accounts, such as the Mind/Body problem and mental causation. But serious concerns still arise for the view that mental state terms, particularly propositional attitude terms, work to license inferences about possible behavior. While the view has some initial plausibility for third-person attributions, the plausibility is drastically diminished for the first person-case. I may attribute beliefs and desires to you because they are consistent with your observable behavior, but it is seems strange to say I come to know my own beliefs and desires by observing my own behavior, or that I aim to license

¹⁵ Armstrong, D. M. (2004). *Truth and truthmakers*. Cambridge University Press.

inferences about what I have done and am likely to do when I tell you what I want. A simple Rylean non-descriptivist view also seems vulnerable to the Frege-Geach objection (Geach 1960), or what we might call the Embedding Problem. Even if we have correctly identified the function of mental-state attributions (as licensing inferences), to identify the function—or to say what we are *doing* in attributing a mental state is not to say what the relevant sentences *mean*. For mental state attributions may also appear in embedded clauses, e.g. “If she thinks the lake is frozen, she hasn’t been watching the weather lately”, which (being embedded in a hypothetical) doesn’t license us to infer anything about how she will behave. A more sophisticated non-descriptivist approach to mental state talk must provide a response to this classic objection, and, in addition to giving an account of the function of mental state talk, give an account of the meanings of our mentalistic terms that remains constant, even in cases where they are not being used to serve that function. I will return in Chapter 4 to discuss this problem.

In the pursuit of challenging the latent Cartesianism within our outlook on the mind, Ryle showed that quite often we appeal to ordinary behaviors when we are discussing beliefs, desires, intentions, or feelings – whether those of ourselves or of others. This discursive practice of settling disputes, answering questions, and justifying claims by making reference to behaviors counts as a mark against the metaphysical and epistemological assumptions of Cartesianism – that mental states are essentially private entities, which we come in contact with through some ‘inner’ sense that privileges the first-person perspective over the third-person perspective. It is obvious to competent language users that if Skyler chooses

vanilla over chocolate and strawberry it's because she *prefers* it, that Walt's unwillingness to enter the house when the Great Dane is present, along with his sudden stammer when talking and flushed face mean that he is *afraid* of the dog, and that Jesse's joining the local gym and buying a new pair of sneakers are evidence of his *intention* to get fit. If mental states were anything like our Cartesian commitments suggest that they are these quite mundane practices of justification would be rendered mysterious. The dispositional profiles of behaviors should not be understood as *referents* of the mental predicates, but instead justificatory criteria for evaluating the aptness of certain mental attributions. Ryle's negative project, to refute Cartesian assumptions underpinning the way many philosophers reason about the mind rendered a certain accounts of mental states untenable. This includes accounts ranging from substance dualism to those physicalist accounts that identify mental states with brain states. If a theory casts the referents of mental terms/concepts to be *internal* to the subject and *hidden* from ordinary observation, it will fall victim to Ryle's objection.

From this discussion of Ryle I take the following lessons: descriptivism is still the default assumption in philosophy of mind today, framing the questions we ask and constraining the types of answers we can give. In one form or another the main positions in philosophy of mind still assert the metaphysical commitment to mental states/processes conceived of as *internal* to the subject, and the epistemological commitment to the *privileged* and *private* way we access and come to know those states/processes (and the corresponding asymmetry to how we access and come to know the mental states of others). Excavating and

challenging these assumptions creates a promising path towards a new understanding of some very old problems.

As mentioned above, I will incorporate Ryle's insight that disjunctive behavioral dispositions are crucial in our evaluation of propositional attitude attributions, while providing a more rigorous framework by which we can stave off both the erroneous reading of the view as problematically behaviorist and the embedding problem that arises for all accounts that give function-driven analysis of a given domain of discourse. I will now turn to an exposition of key sections in Sellars' *Empiricism and the Philosophy of Mind*, to articulate the continued early interest in understanding the function of mental discourse.

§2.2 Sellars and the Mind

Much like Ryle, Sellars had an overarching interest in meta-philosophical issues, apparent in all of the topics that he wrote about. From his earliest publications, Sellars was concerned with how to understand philosophical enterprises in relation to those of science. According to Sellars' view, philosophical inquiries are continuous with, but distinct from, scientific inquiries. The central aim of philosophical inquiry is to make explicit the implicit relations, commitments, and implications of our theoretical and everyday practices. He took philosophy to be a method of becoming reflective and knowledgeable about topics of which one ordinarily only has an immediate and non-reflective understanding. There's a striking resemblance to Ryle's take on philosophical inquiry or what he called

logical cartography, the mapping of conceptual territory that gives us an objective and thorough look at matters to which we previously had only subjective and partial access. Both thinkers took the primary task of philosophy to be articulating commonplace concepts in a systematic way in an attempt to become clearer about the interrelated connections and implications, as well as the contradictions and confusions, which lie dormant there. Also, in Sellars, we find a focus on the plurality of our conceptual/linguistic practices, and an urging to not fall into the trap of thinking of language as serving one purpose. In *Empiricism and the Philosophy of Mind*, Sellars remarks,

"Analysis" no longer connotes the definition of terms, but rather the clarification of the logical structure -- in the broadest sense -- of discourse, and discourse no longer appears as one plane parallel to another, but as a tangle of intersecting dimensions whose relations with one another and with extra-linguistic fact conform to no single or simple pattern. (EPM, 40)

Not long after Ryle's *Concept of Mind* had made its impact, Sellars brought this methodological sensibility to the area of philosophy of mind, advancing a novel nondescriptivist theory of mental talk, particularly of the role of 'appearances' talk in *Empiricism and the Philosophy of Mind*. In this section, I will describe the view and consider its strengths and weaknesses.

Sellars begins with the epistemological goal of criticizing what he calls *the myth of the given*, denying the idea that there is a realm of concepts that derive their meaning solely from an encounter with a special kind of object of experience. The main target here is a Foundationalist theory of knowledge, that our capacity for knowing things about the external world is founded in a more basic ability to know features of our own experience. Such features, according to the broadest

notion of Foundationalism, are "given" to us in experience – and for that reason they are incorrigible or infallible in a way that knowledge gained from perception or testimony could never be. Sellars takes issue with the attempt to legitimize empiricism via a Foundationalist theory of knowledge, denying that there must exist some realm of *given* knowledge in order to rescue our intuition that knowledge of the world is possible. He realizes, though, that he cannot merely criticize the widely popular view, he must also offer an alternative theory to serve as a replacement for the traditional route to defending empiricism¹⁶. In doing so, Sellars advances an elaborate inferentialist theory about the meaning of mental predicates, which explains the *goodness* or *justification* of claims about our own mind in terms of the rules that connect and govern application of such predicates. Claims of knowledge of our own mental states might not be *incorrigible* or *infallible* in the way Foundationalism initially claimed, but they do enjoy a heightened security, compared to other sorts of claims based on, say, perception or testimony, because of the conceptual way we arrive at mental state attributions. In this analysis, we find a similar strategy as noted in Ryle, one that challenges the descriptivist assumption about how mental predicates work, while also drawing attention to normative aspects of the discourse,

The essential point is that in characterizing an episode or a state as that of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says (EPM: 36)

¹⁶ deVries, Willem, "Wilfrid Sellars", The Stanford Encyclopedia of Philosophy (Fall 2015 Edition), Edward N. Zalta (ed.)

The motivation to resist the Foundationalist picture of knowledge by offering a less descriptive, if not *nondescriptive*, treatment of the relevant discourse leads Sellars, later in EPM, to look to the logic of different sorts of mental predicates to see how they function in ordinary language. According to this thread of Sellars' account, talk of *seemings* or *appearances* shouldn't be taken as attempting to describe some inner states. Instead, it has the function of enabling the speaker to withhold world-regarding commitment. The mythical tie-seller, John, having noticed the tricks of artificial light, learns (on the basis of his mistakes) to shift from saying "that tie is green" to "that tie looks green to me", and thereby to withhold from endorsing the claim that it is green:

Now the suggestion I wish to make is, in its simplest terms, that the statement 'X looks green to Jones' differs from 'Jones sees that x is green' in that whereas the latter both ascribes a propositional claim to Jones' experience *and endorses it*, the former ascribes the claim but does not endorse it. (1956/1997, 41)¹⁷

Withholding commitment, on the Sellarsian view, enables us to make the shift from world-talk to mind-talk, particularly in the case of talk about how things *appear*. This might seem to hold little promise for being generalized, as a way to get us up to ascriptions of other mental states. Particularly in the case of belief, for example, it seems that if one is withholding commitment about whether or not P, one should no longer say one believes that P.¹⁸ Sellars, however, seemed to derive his view

¹⁷ He does, however, allow that such claims *also* function as reports: "when I say 'X looks green to me now' I am *reporting* the fact that my experience is, so to speak, intrinsically, *as an experience*, indistinguishable from a veridical one of seeing that x is green" (1956, 41).

¹⁸ Thanks to Katalin Farkas for making this point.

from Husserl's notion that the 'bracketing' involved in phenomenological reduction could enable us to shift from world-talk to phenomenological talk (Thomasson 2005). And if we take it back to the original sense and use of Husserlian bracketing, the idea seems far more plausible. Husserl took the method of bracketing—suspending the assumptions of the 'natural stance' about how the world is, and even suspending the assumption that there is a mind-external world at all—to be the route into talk of the phenomenal realm. But, as Husserl was always careful to emphasize, bracketing cannot be understood as placing those assumptions in *doubt* or coming to hold a different attitude towards them (so we don't shift from belief to doubt) but rather of simply placing the assumption to the side, putting it 'out of action'. So understood, the method of bracketing may after all give us a story about how we can come to know our mental states that doesn't appeal to anything like inner observation, or quasi-scientific positing. On this view, (coherent with the observations of transparency theories), we acquire knowledge of our own mental states not by gazing inward, but rather by attending to features of the world around us. We begin from normal 'use' of our experience to acquire information about the world around us, but then 'bracket' the question of whether the world really is that way. On this view, the rules of use that entitle us to introduce belief talk, for example, entitle me to make a reductive transformation, that might, for example, take me from the judgment "That tie is green," to bracket the question of whether that is true, to shift to "I believe that that tie is green," and come to knowledge of my belief. I can then undertake a further, hypostatizing transformation to speak of mental states explicitly (using noun terms), saying "I

have a belief that the tie is green”—and via this route come to explicitly refer to mental states.

The Husserlian picture gives us the basis for a neat non-introspective story about how we come to be able to speak of the ‘phenomenological realm’, but it doesn’t seem to answer the functional question with which we began: why would we want to have a language that involved not just world-oriented terms but mental state terms? What role do ordinary mental state ascriptions play in our language? And while this commitment-withholding function might be reasonably extended to belief-talk, in the sense that sometimes we might say “I believe John is a decent accountant” when referring his services to you in order to make the stronger claim that “John is a decent accountant,” Sellars’ account of the ‘hedging’ function won’t be generalizable to other types of mental states, such as desire, intention, or doubt. And while Husserl employed talk of the phenomenological with epistemic goals in mind, our everyday talk of our mental states could hardly be attributed such grand philosophical ambitions—and our talk of the mental states of others couldn’t be reached through bracketing at all. For these reasons, Sellars’ take on seemings talk is instructive in the pluralistic nature of how mental discourse may function, but not rigorous enough to be advanced as a thoroughgoing treatment of propositional attitude talk overall.

§1.3 Wittgenstein and the Mind

Wittgenstein was acutely aware of how the descriptivist assumption about the uniformity of language is naïve at best and contaminating at worst. For instance, in the *Philosophical Investigations* (1958), Wittgenstein comments,

Think how many different kinds of thing are called “descriptions”: description of a body’s position by means of its coordinates; description of a facial expression; description of a sensation of touch; of a mood... What we call “*descriptions*” are instruments for particular uses¹⁹. (#24, #291).

This echoes a concern of Ryle’s discussed above, and Wittgenstein asserts that,

The paradox disappears only if we make a radical break with the idea that language always functions in one way, always serve the same purpose: to convey thoughts - - which may be about houses, pains, good and evil, or anything you please. (1958, #304)

In that spirit, Wittgenstein at least tentatively advances a view of mental discourse as functioning in an expressive, as opposed to descriptive, way. The first step towards this thesis is a critique of the commonly held picture of mental states as private. We can run through an example to illustrate this point. According to the common view, each mental state term stands for some internal,

¹⁹ Wittgenstein, L. (1953/58). *Philosophical Investigations* Oxford. Eng.: Blackwell.

privately accessible mental state of a subject. "Pain," for instance, refers to a range of unpleasant phenomenologically introspected states. Feeling pain (being introspectively aware of a certain phenomenologically unpleasant state) immediately justifies my being in pain and my first-person attribution of 'pain.' My attributions of 'pain' to others must somehow derive their justification from extension from the first-person case, since the term's meaning is fixed by the phenomenology of the relevant state. But Wittgenstein worries about the puzzle this interpretation generates – if my understanding of the term 'pain' depends on my introspected knowledge of my own sensations, how can I extrapolate to the third-person case and know that a person wailing and wincing before me is having an experience with *that* quality? If the standard story were true, then my justification for third-person attribution of mental terms would always be terribly weak since the central evidence for mentalistic attributions are barred from me in those cases. The puzzles don't exist only on the linguistic level, it remains a metaphysical mystery how states brought about by physical causes, realized in physical systems, and bringing about physical effects may, in principle, elude a physical analysis due to their essentially privately introspectible phenomenology. We started with the incontrovertible fact that I know my own pains more intimately than I know yours, and ended up building this simple observation into a view about the mind that renders the language and metaphysics of thought impossible. Wittgenstein's underlying nondescriptivism about mentalistic language surfaces explicitly in his critique of the standard story, he says

Perhaps the word "describe" tricks us here. I say "I describe my state of mind" and "I describe my room." You need to call to mind

the differences between language games. (PI 290)

The 'different games' Wittgenstein has in mind are the other *nondescriptive* ways in which language can be employed. This leads to an expressivism towards certain mentalistic discourse. For example,

Here is one possibility: words are connected with the primitive, the natural, expression of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations, and later, sentences. They teach the child new pain behavior. "So you are saying that the word 'pain' really means crying?" – On the contrary: the verbal expression of pain replaces crying and does not describe it. (1958, 244)

Recently a neo-expressivism about language involving mental concepts has been advanced, most notably by Bar-On (2004) and Finklestein (2003)²⁰, that treats utterances like "I'm in pain" or "I want a drink of water" as expressions of underlying mental states. These views attempt to distance themselves from what Bar-On calls Simple Expressivism, a view that she attributes to Wittgenstein as indicated in the above passage. The problem with Simple Expressivism, in contrast to the neo-expressivism Bar-On and Finklestein defend, is that Wittgenstein seems to have taken linguistic expressions of mental states to be no different in kind than non-linguistic expressions of mental states, such as crying out when one is in pain, or reaching for a glass of water when one is thirsty. Simple expressivism, as (perhaps) advanced by Wittgenstein, can be applied not only to sensation terms like 'pain,' but also to propositional attitude terms. "I

²⁰ Bar-On, D., 2004, *Speaking My Mind: Expression and Self-Knowledge*, Oxford: Oxford University Press.; Finkelstein, D., 2003, *Expression and the Inner*, Cambridge, MA: Harvard University Press.

believe the table will fit in the den” is a cautious statement about furniture and rooms – if I asserted, “the table will fit in the den,” I have communicated a confidence in the certainty of that proposition. But by asserting, “*I believe* the table will fit in the den,” I have softened the commitment to the proposition, subtly expressing some doubt about its certainty. In a sense, the Simple Expressivist view of belief is consistent with Sellars’ commitment-withholding function of *seemings* talk.

The view attributed to Wittgenstein by Bar-On as Simple Expressivism shares a lot in common with emotivism in ethics. “I’m in pain because I stubbed my toe,” means something like “Ouch, toe!” while “I want a glass of water” might be interpreted as “Yay, water!” This view, of course, will run into the embedding problem shared by the Rylean account of mental state terms as inference tickets and the Sellarsian understanding of seeming-talk as commitment-withholding. And while the Rylean view makes sense in the third-person, but seems implausible for first person attribution; Simple Expressivism shares with the Sellarsian story will strength in the first-person case, but leaves mysterious how third-person attributions are meant to function. Just as I cannot withhold commitment to propositions you might assert, I also cannot express verbally states of affairs that might have been expressed more naturally with yawns or yelps by you.

§ 2.4 Deserting Descriptivism

The historical views I've discussed here do not, independently or combined, constitute a thoroughgoing theory of how mental discourse functions. What these views share in common is a willingness to uncover hidden assumptions about the semantics of mental discourse, and an insistence that when we take all language to serve a tracking function we end up creating seemingly deep problems that will only truly be resolved by better understanding the logic of the discourse. The motivations that led Ryle, Sellars, and Wittgenstein to depart from descriptivism decades ago are still present in our contemporary debates. And while the respective views advanced by each fell by the philosophical wayside, rejected not only for what were taken to be devastating objections but also for contingent historical details about what was fashionable at the time²¹, I believe there is still a good deal of value to be taken from the nondescriptivist strategies they pioneered. The category mistake inherent in both dualistic and physicalist theories, is still found in the way we approach studying mental phenomena today – and are still as problematic as ever.

§2.5 Chapter Conclusion

Using the insights of earlier views as the starting point, my project is to distill the insights of these views, along with more recent nondescriptivist threads in

²¹ Say something here about Logical Positivism's downfall, Ryle's perceived connection to that view (logical behaviorism), and Sellars' view swift transformation into a type of proto/functionalism.

philosophy of mind (such as Dorit Bar-On (2004) and David Finklestein's (2012) expressivist work²²), and unify them into a single view that makes explicit the shared meta-linguistic commitments, strengthens the view against the worst objections that have been brought against nondescriptivist accounts, and draws out any metaphysical and epistemological implications. I see the positive contribution of this work as synthesizing consonant research from philosophy and psychology about the way mentalistic discourse functions into a broad account that is able to reinterpret classic problems in philosophy of mind in a new light. Although the view I defend here is primarily *meta-philosophical*, in the sense that it criticizes underlying assumptions prevalent in philosophical debates about the mind and asks what the landscape of the topic would look like if we began with *different* assumptions, I don't see the argument as merely negative. The way that the folk, philosophers, psychologists, and neuroscientists conceive of mentalistic discourse, and subsequently the mind, has important practical implications for the way the mind and brain are studied by experts.

The troubles that arise directly from the descriptivist assumption about how mental discourse works might tempt us to look at the alternative. What would the landscape of philosophy of mind look like if it weren't concluded from the outset that attributions of mental states function to describe a class of entities or features of the world? Of course, this questions isn't new. In Chapter 2 I reviewed some of the reasons that led theorists like Ryle, Sellars, and Wittgenstein to go

²² Bar-On, D. (2004). *Speaking My Mind: Expression and Self-Knowledge: Expression and Self-Knowledge*. Clarendon Press.; Finkelstein, D. H. (2012). From Transparency to Expressivism1. *Rethinking Epistemology*, 2, 101.

nondescriptive about certain aspects of mentalistic discourse. A recognition of the plurality of ways language functions, coupled with a reticence to go down the descriptivist path knowing full well where it leads us, prompted some twentieth century philosophers to go nondescriptivist about a number of problematic areas of discourse, including the mental. Yet, as illustrated in Chapter 2, these views were not without their own problems. Ryle's work on the importance of disjunctive dispositions for analyzing mentalistic language was wrongly labeled behaviorist, many read the view as claiming an identity between terms like 'belief,' 'desire,' and 'intention,' and certain disjunctive dispositions instead of understanding the dispositions as criteria for *evaluating* and *settling disputes* regarding mental state attributions. But even if we read it charitably and reject the standard behaviorist label, Ryle's view is not an adequate account of propositional attitude attribution. We may invoke dispositions to behave when evaluating attributions of mental states in the case of *other people*, but in the first-person case, Ryle's view falls flat. And more worryingly, if we try to identify the meaning of mental state terms with the purported function of those terms, we run head first into the embedding problem. Sellars and Wittgenstein's respective nondescriptive projects face similar objections. Whether we're using 'appearance' or 'seemings' talk to withhold commitment as Sellars claims, or using mental terms as verbal expressions on par with more basic, nonverbal, expressions like crying to express sadness, or wincing to express pain, as Wittgenstein argues, it is implausible in the case of third-person attributions and still incapable of sufficiently solving the embedding problem. Nondescriptivism needs some further sophistication in order to stand up to these

objections. Now that I've explored the genealogical beginnings of mentalistic descriptivism, in the next chapter I develop a function-first account of propositional attitude discourse, which builds off older views while incorporating relevant social and developmental psychology and contemporary philosophy to demonstrate the normative role that talking about beliefs, desires, and intentions plays in our lives.

CHAPTER 3: NONDESCRIPTIVISM AND THE PROPOSITIONAL ATTITUDES

Overview

Most of us come to study philosophy of mind without commitment or conviction towards the 'true nature' of the subject matter. We enter the academic study of the mind as undergrads with only our pretheoretical intuitions about what types of questions we'll encounter and topics we'll explore. Typically, the classic problems are explained to us in a certain way; in simple terms at first, but progressively growing more complex and nuanced as we advance in our careers. There are a number of core problems, which splinter off into smaller questions and sub-debates - for instance, the mind/body problem, causation, and views of interaction; the hard problem of consciousness in its various articulations, the knowledge argument, zombies, inverted spectra; the problem of other minds and the debate over asymmetry between self knowledge and knowledge of other minds; issues of representation/anti-representation, intentionality and the internalism/externalism debate about mental content, and other questions of content determination and character. Stepping back, we can notice a theme in how these topics are unfolded for us. Our attention is drawn to the apparent differences between mental states and physical states, or mental properties and physical properties, self-knowledge vs. knowledge of others, and we're asked to consider what type of *thing* mental entities could be, and what would have to be true of the world in order to bridge the gaps mental states allegedly create in our explanations of the world. We're

invited into what often are centuries old debates about the nature of the mental, causal interaction, and self knowledge, and learn all the positions one can adopt toward these philosophical puzzles, the problems with those positions, and the rejoinders to those problems. The debates we encounter are largely metaphysical and epistemological, and rarely focus specifically on pragmatic questions about the nature of mental discourse. At the outset of our learning, we don't have the philosophical sophistication to get a little distance and notice that assumptions are being made about the way mental discourse functions. The standard stories about the mind, although diverse in formulation and response, share a common pragmatic commitment – descriptivism.

Descriptivism is a view about how a specific domain of discourse functions, specifically that language in a certain domain's *purpose* is to *describe* the world. In Price's terms, descriptivism is a commitment to a view of language as *e-representational*, where noun terms covary 1:1 with things in the world, serving to reliably *track* and *report* on those things. My goal here is to persuade you that the underlying commitment to descriptivism that frames many of these classic debates in philosophy of mind is suspect. First, it's naïve to assume that *all* language involving apparent reference to the mind/mental states aims to *describe* a special type of entity. As we have seen in Chapter 2, Ryle, Sellars, and Wittgenstein made this point decades ago. But more importantly, expecting functional uniformity in mental discourse has created more problems than it has solved. For this reason, challenging that assumption is wise on the methodological level. But there also exists evidence from cognitive and developmental psychology that supports the

claim that the language of propositional attitudes primarily serves a normative, as opposed to descriptive, function. I will argue here that the combination of methodological and empirical reasons for accepting propositional attitude nondescriptivism render the view both plausible and favorable compared to the alternatives.

There are two main strains of descriptivism prevalent in philosophy of mind and cognitive science, direct observation and indirect observation views. I will briefly describe these two widely popular and influential views on the function of mental discourse in order to articulate how they give rise to persistent problems in philosophy of mind. This will allow me to better demonstrate, once I've argued for the Intelligibility View, how going nondescriptivist uniquely positions us to avoid the well known problems that arise from a descriptivist understanding of propositional attitude discourse.

§3.1 Direct Observation Descriptivism

As discussed in Chapter 2, the historically most influential view on how mental discourse functions is that it aims to describe a certain kind of entity that we observe by means of introspection, this is the semantic claim inherent in the "official doctrine" identified by Tanney. Just as vision allows us to track the squirrel as it scurries from branch to branch, and hearing enables us to track a train as it approaches and then trails into the distance, the Direct Observation view tells us that there is a distinct faculty for tracking mental entities, a type of 'inner sense.'

While this view is present in philosophy as far back as Plato and Aristotle, this view can be thought of as broadly Cartesian since Descartes' systematic discussion of the relationship between minds and bodies, and how we come to know these things, in the *Meditations* gave the view its most famous formulation. For an explicit account of such a faculty, Shoemaker distinguishes between two versions of the perceptual model of introspection, or what he calls 'inner sense' in his 1994 article *Self- Knowledge and Inner Sense*²³. The first version is the *object perception model*, in which the word *object* designates particular things and excludes facts. The second version is the *broad perceptual model* in which we have access to both *things* and to *facts/states of affairs*. Both models ascribe to perception the following feature,

...in perception we have access to things or states of affairs that exist independently of their being perceived and independently of there being any means of perceiving them." (1994, 204)

The Direct Observation view has many contemporary proponents and is not at all limited to Dualism.²⁴ Any view that broadly construes mental states as internal to the subject and observable via some inner sense counts as a direct observation view. The Direct Observation treats discourse about mental phenomena on analogy with discourse about physical phenomena that we come to know via our sensory faculties—assuming that have a similar tracking and describing function.

²³ Shoemaker, S. (1994). Self-Knowledge and "Inner Sense": Lecture I: The Object Perception Model. *Philosophy and Phenomenological Research*, 54(2), 249-269.

²⁴ Advocates of the Direct Observation represent a variety of views about the nature of the mental: Gertler's acquaintance view of self-knowledge, Fumerton's defense of a type of dualism, Jackson's physicalist functionalism

As alluded to in Shoemaker's distinction, Direct Observation views characterize the language of propositional attitudes, and mentalistic language more broadly construed, as performing the function of referring to independently existing objects of awareness.

This assumption about function leads to both ontological and epistemic consequences. The ontological consequences stem from understanding mental discourse as running functionally parallel to discourse involving objects of perception, so that we end up relying too heavily on the vision analogy, and look for a distinctive faculty of introspection by which these objects are known. In addition to the original mind/body problem that rests on the descriptivist assumption, and the related problem of mental causation that naturally arises when we wonder how these mental things can (non-redundantly) figure in causal explanation, we now add the new metaphysical problem of accounting for the faculty of introspection. These ontological problems will arise from any descriptivist understanding of mental discourse.

The descriptivism characteristic of the Direct Observation view also leads to distinctive epistemic concerns. First, the view that the strongest and least defeasible justification for knowledge of mental states results from a kind of inner observation analogous to external perception, leads to a version of the problem of other minds. For mental states, according to the Direct Observation view, are privately observable from one's first person perspective (using introspection), and never publically observable in the way physical objects are. Of course, an advocate of the Direct Observation view may accept that we can have *indirect*

behavioral/verbal evidence of mental states (whether of ourselves or others), which can lead to certain justified (and possibly true) beliefs about the mental states of others. Nonetheless, if the only evidence we have is so indirect and easily overturned, it becomes mysterious how we could come to know another person's mental states despite the fact that we seem to understand each other quite well in our ordinary social interactions. In everyday discourse about what we think, we attribute knowledge, belief, desire, and depression easily to others and ourselves. But if such terms refer to states belonging to another person's *inner life*, then we have trouble giving sufficient justification for these attributions, and can never give anything but very fallible and indirect justifications, compared to the immediate and incorrigible evidence we have of our own inner lives. Only the subject of the relevant mental event would be able to let us know if the term had been correctly or incorrectly applied. And that person would never be able to be contradicted or corrected about the states of her own mind. But this again is in conflict with our everyday practices of attributing mental states, since a person's sincere claim to believe or want or doubt what she says she does is only *one* type of evidence we take into account when settling such matters. If each of us had solely personal access to the strongest type of justification in mental state attribution, our discursive practices begin to appear misguided, if not impossible.

But the issues don't end with third-person attribution; similar problems arise for attributing mental states to oneself. Since *directly observing* your own mental states will deliver the most immediate and dependable justification for the belief that you are in such-and-such mental state, the indirect evidence can never trump

the direct introspective evidence. Just as directly seeing a squirrel is better evidence that there is a squirrel in front of you than seeing squirrel tracks, directly introspecting one's own anger will be better evidence for the presence of an anger-state than observing clenched fists and a flushed face. But this means that a subject can't be wrong when it comes to her own mind, whatever evidence a third party may have that seems to contradict that the subject is angry will always be trumped by the subject's own introspective justification. It does, however, seem that we *are* sometimes wrong about our own minds, and can have our own judgments about what we're experiencing overturned by an attentive friend or psychologist. Moreover, the (alleged) faculty of introspection has no obvious dedicated organ in the way that other sensory modalities do. 'Dedicated organ' is to be understood here as the physically and functionally individuated system that gives rise to perceptions of a certain kind. Sensory systems can be localized or dispersed in the body. For example, the visual and auditory systems are rather localized, while those of proprioception and interoception are more diffuse. Objects are visually processed in our bodies by the eyes and corresponding neural areas, hearing by the ears, and olfaction the taste buds. But there is no candidate for a system within the body dedicated to detecting and processing mental states. While we've amassed a great deal of research regarded proposed 'neural correlates of consciousness,' for instance suggesting that heightened activity in the dorsal lateral prefrontal cortex is implicated in conscious experience, there is little research to suggest that there is any structurally and functionally dedicated system discretely for detecting one's own mental states. Accounting for the faculty by

which we come to know our mental states arises for any descriptivist view, direct or indirect. But it's more pointedly a problem for Direct Observation, since the account of the function of mental discourse offered by these theories relies so heavily on the analogy between outward senses and the alleged 'inner' sense.

Another problem that Direct Observation views face is that they run afoul of the Transparency Thesis, the claim that it is impossible to attend to properties of experience itself, and that when we try to do so we are attending to the objects/properties of which the experience is about. A number of philosophers have argued that knowledge of our own minds is *transparent*. Ordinarily when you're driving, you look through the windshield to the world outside. You see traffic, road signs, and the scenes streaming past at your sides. Analogously, it has been said that we normally 'see through' our mental states to the things that they target. When you are aware that you want some French fries, it's the merits of salty hot potatoes of which you are aware. When you realize that you doubt the electability of a candidate, you are focused on the politician's lack of experience or likeability. An advocate of the Transparency Thesis draws our attention to how these experiences are *of* 'objects' in the world, and how we are made aware of their properties through our mental states, just like we are made aware of stoplights and bumper stickers through our windshields. Transparency theorists hold that experience is always transparent in this way; I am able to introspect properties intrinsic to the objects of my experience, never properties that belong to 'experiencing' itself. There are plenty of examples detailing transparency in the literature: Moore (1925), Tye (2002), Harman (1990). One clear, often cited

description is given by Gilbert Harman²⁵,

When Eloise sees a tree before her, the colors she experiences are all experienced as features of the tree and its surroundings. None of them are experienced as intrinsic features of her experience. Nor does she experience any features of anything as intrinsic features of her experiences. And that is true of you too. There is nothing special about Eloise's visual experience. When you see a tree, you do not experience any features as intrinsic features of your experience. Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict you will find that the only features there to turn your attention to will be features of the presented tree. (1990, p. 667)

This example asks us to try to turn our attention to some specific feature within our experience, or to attend to the experience in a certain way. It is important to say a word about what attending to an experience means. When I have a visual experience of a tree, I am directly aware of the properties of the tree, perhaps the tree's height, color and shape. Now, if I were to concurrently 'attend' to my visual experience, my attention refocuses, it is allegedly the *experience* of myself seeing the tree that I am attending to, instead of the tree itself. The transparency thesis tells us that this shift in focus cannot reveal any *new* properties within the experience. A proponent of Direct Observation must claim that experience is not transparent in the way described above. We can be perceptually aware of the world without introspection, that's why we are asked to 'turn our attention to intrinsic features of the perceptual experience' in the example. But turning our

²⁵ Gilbert Harman, "The Intrinsic Quality of Experience," in *Philosophical Perspectives*, edited by James Tomberlin, Vol. 4 (Atascadero, CA: Ridgeview Publishing Co., 1990), 52-79; reprinted in *The Nature of Consciousness*, edited by Ned Block, Owen Flanagan, and Guven Guzeldere (Cambridge, MA: MIT Press, 1997), 663-75.

attention to the experience, as opposed to the tree, doesn't give us any new information that wasn't already contained in the original experience (although focusing intently may *amplify* the qualities of which we're aware, they are qualities of the perceived objects, not experience itself). Given transparency, introspection must be meant to do some *other* job besides tell us what we are perceiving, provide us with some *other* type of information about the 'inner' world and experience itself. Anyone who wishes to defend a view of introspection that is quasi-perceptual must explain why perception is transparent while introspection is not. We are owed an account of the intrinsic features of experience that we gain access to by introspection, and until such an account is forthcoming, we have no reason to accept that there is distinct mental faculty providing such access. The model of introspection necessitated by a Direct Observation view of mentalistic discourse is rendered redundant by the Transparency Thesis.

For all of these reasons, the type of descriptivism typified by the Direct Observation view deserves to be held in high scrutiny. If there's an understanding of mental discourse that is able to avoid the epistemological and metaphysical problems pushed on us by this form of descriptivism, it is in our best interest to pursue it.

§3.2 Indirect Observation Descriptivism - Mind Reading

Moving away from the view that mental states are directly observed entities, another general approach to understanding propositional attitude attribution is that of *mind reading*. Adopting the terminology of Nichols and Stich (2003) 'mind

reading' is the branch of folk-psychological theorizing that deals with the following²⁶:

- (a) the ability to predict human behavior
- (b) the ability to attribute mental states to oneself and others
- (c) the ability to explain behavior in terms of possession of mental states

Taking these three claims together, the Mind Reading view tells us that mental state attribution is a cluster of capacities that allows us, by attributing mental states to others and ourselves, to predict and explain human behavior. According to this view, the development of human social understanding mirrors the scientific method in important ways. Children form social theories based on environmental observation, and revise and update that theory as new observations are made. On this view, children begin to postulate mental states at a key stage of development, in order to better predict and explain the seemingly confusing behavior of people with whom they interact. Finally, the child comes to understand her own behavior in terms of postulated mental states like belief, desire, intention, and doubt. On this view, mental states are never directly observed, but akin to electrons, neutrinos, and gravitational forces within scientific theories are postulated in order to make the best sense of entities and behavior that *is* observed. Mind Reading counts as

²⁶ The three capacities listed are the *central* capacities that fall under 'mind reading'. There are related but distinct cognitive capacities in addition to these such as the ability to predict and explain the mental states of oneself and others (as opposed to behaviors), the ability to evaluate the actions and minds of others in terms of the possession of mental states, and the ability to attribute mental states to non-human animals and non-animal systems. These sub-capacities are generally taken to be derivative of the central three.

a descriptivist view, since according to the view mental discourse still functions to describe certain entities in the world; it's just that the analogy has changed. Instead of mental discourse working like perceptual discourse, to track and describe entities we come directly in contact with by some bodily faculty, now mental discourse is said to work more like the theoretical terminology in science, postulating indirectly observed entities which fill the gaps in our theories about the world.

Interestingly, alongside early attempts at nondescriptivism about certain mental vocabulary, the Mind Reading view also has roots in the work of Wilfred Sellars. In *Empiricism and the Philosophy of Mind*, Sellars proposed a view of this type as an explanation of how mental discourse might have first been introduced to human language. Sellars has us imagine a society, complex in its linguistic, conceptual, and social dimensions, but lacking any reference to mental states – he calls this hypothetical tribe "our Rylean ancestors²⁷." Prior to the introduction of mentalistic vocabulary, we're told that this society's

...total expressive power is very great. For it makes subtle use not only of the elementary logical operations of conjunction, disjunction, negation, and quantification, but especially of the subjunctive conditional. (1956, 48)

In his view, just as we might come to posit unobservable entities, like electrons and centers of gravity, in order to maximize the explanatory and predictive power of our theories, so might the mythical tribe come to introduce psychological terms

²⁷ Sellars, W. (1956). Empiricism and the Philosophy of Mind. *Minnesota studies in the philosophy of science*, 1(19), 253-329. p. 178 section 48-50)

for entities posited to help explain and predict the dispositional regularities observed in the behaviors of group members. Sellars' asks the question,

'What resources would have to be added to the Rylean language of these talking animals in order that *think*, *observe*, and have *feelings* and *sensations* as we use these terms?' And 'how could the addition of these resources be construed as reasonable?' (1956, 49)

In answering these questions, Sellars argues that our Rylean ancestors would be in a position to introduce such language at the point that they are able to assess the patterns of verbal behavior in semantic terms. With such semantic resources, it is reasonable for the fictional tribe to talk about beliefs, desires, and intentions - the characteristic intentionality of such terms is rooted, Sellars argues, in the ability to move from specific verbal performances to general semantic principles. Based on this understanding of mentalistic terms as extensions of semantic categories fundamentally rooted in overt, verbal behavior, Sellars explains why a society possessing no mentalistic vocabulary would come to introduce such discourse,

It will not surprise my readers to learn that the second stage in the enrichment of their Rylean language is the addition of theoretical discourse. Thus, we may suppose these language-using animals to elaborate, without methodological sophistication, crude, sketchy, and vague theories to explain why things which are similar in their observable properties differ in their causal properties, and things which are similar in their causal properties differ in their observable properties. (1956, 52)

In Sellars' story, mental discourse is introduced to the Rylean language for *theoretical*, as opposed to *observational*, reasons. The first enrichment necessary to the Rylean language would be that of theoretical discourse, the ability to appeal to entities not observed but methodologically helpful in investigating the world. The

mythical society, as they embarked on scientific exploration, would come to posit unobservable entities, such as electrons and centers of gravity, in order to make sense of the world around them. None of this entails the existence of mental entities, but it is the first condition necessary for moving towards the introduction of concepts that will map up with our contemporary folk psychological predicates. In Sellars' parable, at some point a particularly bright member of the community, Jones, wishes to come up with some explanations and predictions about the dispositional regularities²⁸ he observes his neighbors exhibiting. He wonders why after a few hours of not having anything to eat or engaging in food-acquiring behavior some folks begin to gather and consume food, while others do not. How can the same environmental feature (a certain lapse of time) result in different individual responses: some people searching for and consuming food and other failing to do so? Furthermore, Jones wonders how can *different* environmental features cause people to engage in the *same* behaviors: sometimes his neighbors shed tears when they are reunited with someone they have not seen in a long time, while others shed tears when someone they live closely with passes away²⁹. The answer Jones comes up with is that there must exist unobserved entities *within* these individuals that are the *true causes* of the public behavior, and he calls these *thoughts*. He quickly disseminates this new proto-psychological theory to his community members. Now we are in a position to see what thought-talk allows the Ryleans to do. The theoretical entities 'thoughts' are modeled on public language

²⁸ By 'dispositional regularity' I mean typical behavioral responses of humans to certain environmental and behavioral circumstances.

²⁹ Examples are mine offered to flesh out the more general points made by Sellars in EPM.

in that they are taken to be *silent, private, internal* linguistic episodes. The attribution of mental states, particularly propositional attitudes primarily serve a theoretical, not observational, function for the Ryleans.

Sellars' just-so story about how mentalistic vocabulary might first be introduced by creatures previously lacking such terms was a predecessor to contemporary Mind Reading views that have become increasingly popular in cognitive psychology and philosophy. Arguably, the most popular Mind Reading view is Theory-Theory, which operates on an Indirect Observation model of mentalistic language³⁰. Let's consider the core claims of that view.

One very popular view of how we come to attribute mental states, primarily propositional attitudes, is Theory-Theory, advanced by Gopnik & Meltzoff (1997). According to this view at a certain stage of development, between the years of early childhood and adolescence, humans begin to acquire and use a body of internally represented concepts that allow for prediction and explanation of human behavior. These concepts include the propositional attitudes such as desire, doubt, belief, and intention but also encompass a range of other mentalistic concepts such as moods and traits. This theory takes development of cognitive capacities related to mental state attribution to be *innate*, as opposed to learned.

³⁰ Recently Simulation Theory has gained popularity as an alternative to Theory-Theory. I won't go into detail of that view here, but defenses of the view can be found in Gordon, R. M. (1986). Folk psychology as simulation. *Mind & Language*, 1(2), 158-171, Heal, J. "Replication and functionalism." (1986), and Goldman, A. I. (1992). In defense of the simulation theory. *Mind & Language*, 7(1-2), 104-119.

Describing the mentalistic concepts that develop starting in early childhood, Gopnik says,

These representations and rules are often an interesting combination of the logical and psychological: they are abstract structures, often described in terms of an implicit computational model, but they are also intended to be psychologically real descriptions of how the mind works. The representations and rules may not have any special phenomenological mark, one way or the other, we may not know that we have them, though sometimes we do. They may be deeply influenced by information that comes from other people, but they are not merely conventional and they could function outside of any social community.³¹

The central claim is that these concepts are acquired, stored, and deployed in much the same way scientists generate theories – hypotheses are generated, predictions are made, experimentation takes place, and then hypotheses are modified or replaced as predictions fail. Toddlers and young children, according to this view, are proto-scientists, building theories based on the information they have, and interpreting, revising, and transforming the theories in light of observation and experimentation. Among the concepts they acquire in this process are those of belief, desire, doubt, and intention.

Popular as Indirect Observation views like Theory-Theory have become, various problems arise for the idea that mental state ascriptions function as quasi-scientific posits that aid in explanation and prediction. First, in contrast to the Direct Observation view, insofar as Mind Reading views are committed to the claim that attributing mental states is primarily concerned with explanation and prediction, they have little or no plausibility for the first-person case, whatever their plausibility

³¹ Gopnik, A. (1996). The scientist as child. *Philosophy of science*, 485-514

for third-person attributions. It is only under the most unusual circumstances, situations of breakdown of one sort or another, or psychoanalysis in response to breakdown, that I would come to hypothesize that I may have a fear of drowning as a way of attempting to explain or predict my own behavior of taking the bridge rather than the ferry. The idea that the normal function of discourse about our own mental states is to posit unobserved entities that help explaining and predicting our own behavior seems, to put it mildly, extraordinarily ill-fitting. But even for the third-person case, there are serious disadvantages to considering propositional attitude attribution to be a type of theory. In the case of Theory-Theory, there's a striking disanalogy between scientific theorizing and what we do when we attribute mental states to others. In the case of the latter, there seems to be no prospect of confirming or disconfirming these 'posits', as there normally is for the posits of a good scientific theory that, for example, posits microorganisms, a force of gravity, or electrons to explain observed effects. One may, of course, gather confirming or disconfirming evidence that Lucy is angry by observing her or interacting with her over a period of time, but what parallel is there for confirming or disconfirming the existence of propositional attitudes conceived of as 'theoretical entities,' at large? It remains unclear what such confirmation or disconfirmation would amount to.

There exists a great deal of empirical evidence that we do not need or rely upon propositional attitude ascriptions (at least in normal, everyday situations) in order to predict behavior. Andrews argues that it is only in anomalous cases (where one has no relevant experience, or the behavior was unexpected), where we are also motivated to portray the actor in a positive light, that we come to

explicitly appeal to mental states in order to explain behavior (2012, 110-11). This is a problem for Mind Reading views, which maintain that the purpose of attributing such mental states is fundamentally a task of prediction and explanation. As Andrews puts it,

The process is straightforward: we attribute a specific mental content and attitude to the target, then use a folk psychological theory – or perhaps a mental stimulation, or some combination of theory and simulation – and then generate the prediction. The claim of standard folk psychology is that the richness of our social lives, the ability to anticipate what others are going to do and to understand what they have done, is made possible by this ability to attribute propositional attitudes to others. According to this view, the folk rely on propositional attitudes because they think that propositional attitudes are what cause behavior. (2012, p. 7)

But Andrews points us to a wealth of research from cognitive science that shows that propositional attitude attributions are typically invoked only to explain and predict behavior *in anomalous situations*. If you are walking down the sidewalk and someone is walking towards you directly in your path, you might want to predict what she will do to avoid a collision. But in this case, it's silly to think you must attribute to the person a belief that stepping to the left will clear the path, or the desire not to run straight into you. Or when on the road, we can predict what other drivers will do based on our knowledge of traffic laws, local habits, weather conditions, etc. In such everyday cases, we can assume that the person will step to the left or not turn on red based on past experience and environmental factors, which will generate the quickest and most simple prediction. A theory of mind is not necessary for the majority of the types of explanations and predictions we make about human behavior. Furthermore, children don't acquire mental state language until about age two and a half, and don't pass traditional 'false belief'

tests that require them to contrast what someone thinks with what is really the case until about age four. But they can predict the behavior of others far earlier (Andrews 2012, 24), perhaps using a variety of tools including generalizations from self, consideration of stereotypes, past behavior, and relevant traits. Nonetheless, the evidence that we can and do (in normal circumstances) predict behavior quite well without attributing mental states to ourselves or others, and that we tend to invoke it in explanations *when we are motivated to portray the actor positively* provide important clues that the function of describing people as having certain beliefs, desires, etc. may not be best understood as postulating a hidden entity for explanatory and predictive purposes. Indeed Andrews argues persuasively that the standard 'theory theory' that models folk psychological understanding on scientific explanations (thinking of folk psychology as giving us explanations and predictions based on positing mental states and fitting them into general covering laws) has simply 'imported wholesale' a theory of scientific explanation where it may not belong (2012, 128-32). For after all, theories of scientific explanation are normative theories about how scientific explanation *ought to* proceed; what we are after is a descriptive functional account of what we *are doing* when we make propositional attitude ascriptions.

Perhaps a more crucial concern about accepting either Mind Reading view is that if we genuinely think of these ascriptions as making theoretical-explanatory posits, like Vulcan or phlogiston, or representational models used to predict and explain behavior, then we introduce the possibility that they might turn out to be *failed* posits. Of course, some would embrace this possibility, or even the

conclusion that mental states *are* failed posits³². When we take mental states to be posits or models used to predict and explain, then it follows that if we find better predictors and explainers for our behavior then mental states might turn out to not exist at all. Indeed without a lot of philosophical background, I think we wouldn't even know what to make of the idea that mental states might turn out to not exist, to be merely 'failed posits'. Perhaps it is not quite so hard to make sense of the idea that *others* might turn out to be mere zombies or automata, as Descartes once imagine, this is an idea frequently represented by non-philosophers in stories or films. But there does seem to be at least this much to the idea of Cartesian first-person privilege: that I just don't know what to make of the suggestion that *I* might be wrong about having mental states whatsoever. If even the potential for eliminativism seems misguided, then the Indirect Observation strain of descriptivism may also best be shunned³³.

§3.3 Going Nondescriptive

In the preceding sections, I looked at the two main descriptivist strategies for understanding mental discourse, the Direct and Indirect Observation views. After explaining in broad strokes how each view attempts to make sense of mental state

³² Such as P.M. Churchland, 1981 *Eliminative Materialism and the Propositional Attitudes*, 1988 *Matter and Consciousness*, 1993 *Evaluating our Self Conception*; J. Bickle, 1992 *Revisionary Physicalism*

³³ Dennett's *Intentional Stance* seems to best fit under the Indirect Observation category, since the central claim of the view is that propositional attitudes are attributed because it is *useful* to do so in order to predict and (causally) explain behavior. "All there is to being a true believer is being a system whose behavior is reliably predictable via the intentional strategy, and hence *all there is* to really and truly believing that *p* (for any proposition *p*) is being an intentional system for which *p* occurs as a belief in the best (most predictive) interpretation" (1987, p. 29). For reasons discussed in this section, Dennett's commitment to the attribution of propositional attitudes playing a predictive/explanatory role is problematic.

attributions, I considered objections that each account respectively faces. Direct Observation descriptivism claims to capture ordinary intuitions about how we use mentalistic language, but results in the proliferation of philosophical puzzles about the epistemology and metaphysics of mind – including the mind/body problem, problem of causation/interaction, the problem of other minds, and troubles with accounting for the faculty of introspection. It's also worth noting that any intuitive appeal Direct Observation enjoys could be the result of the pervasiveness of the descriptivist assumption, and not a reflection of the view's theoretical or empirical adequacy. Alternatively, the descriptivist Indirect Observation approach to mental discourse has the benefit of avoiding some of these philosophical puzzles mentioned above, since it rejects the claim that mental states are necessarily private, inner, episodes known by the subject through a privileged mechanism of introspection. However, Indirect Observation views still face problems of causation and interaction, owing us an explanation of how these unobserved theoretical entities fit into explanations of behavior. In addition, this view raises new worries, since it doesn't appear to best capture the findings of cognitive science and introduces the threat of eliminativism. And finally, any descriptivist account of a certain domain of discourse, mental or otherwise, will run into what Huw Price calls the Placement Problem. As he says in *Expressivism for Two Voices*³⁴

Initially, these problems present as ontological or perhaps epistemological issues, within the context of some broad metaphysical or epistemological program: empiricism, say, or physicalism. By the lights of the program in question, some of the things we talk about seem hard to “place”, within the framework

³⁴ Price, H. (2011). *Expressivism for two voices. Pragmatism, Science and Naturalism*. Berlin: Lang.

the program dictates for reality, or for our knowledge of reality. Where are moral facts to be located in the kind of world described by physics? Where is our knowledge of causal necessity to go, if *a posteriori* knowledge is to be grounded on the senses?

Whether attributions of beliefs, desires, intentions, and doubts work like perceptual object language, or instead work like postulations of electrons and neutrinos, then the entities corresponding to those terms are causally efficacious and epistemically substantive in a way that demands elaboration on the part of both the Direct and Indirect Observation theorists. But accounts of mental interaction and self-knowledge are notoriously fraught with difficulties that have proven undissolvable despite the valiant efforts of centuries of philosophical work.

Fortunately, nondescriptivists have made some serious progress in recent decades. In what follows I will argue that a propositional attitude nondescriptivism that is careful in its construction can not only avoid the problems of descriptivism, but can incorporate the insights of earlier nondescriptive attempts while remaining resistant to the troubles I've discussed here. Certain descriptivist problems show up very differently once we understand the function of the relevant discourse. Of the placement problem, Price says,

...the placement problem for moral or causal *facts* rests on a mistaken understanding of the function of moral or causal *language*. Once we note that this language is not in the business of "describing reality"...the placement problem can be seen to rest on a category mistake. (2010, p. 3)

The proposal I offer also make progress on the problems that older versions of mentalistic nondescriptivism faced, by adopting recent strategies I will argue for a

version of propositional attitude nondescriptivism that unifies accounts of first and third person attribution while solving the embedding problem.

§3.4 The Intelligibility View

Let's begin with the functional question: what does belief attribution do for us, if not to describe entities we either directly or indirectly countenance? Recent empirical work may help shed light on this question. In *Do Apes Read Minds? Towards a New Folk Psychology*³⁵, Kristin Andrews argues that it's a mistake to think of folk-psychological explanations on the model of scientific explanation, as serving to explain and predict behavior, appealing to other pragmatic goals instead:

Empirical evidence demonstrates that people offer explanations of their own and others' behavior to fulfill a number of pragmatic goals; we explain behavior to impress other people, to condemn other people, and even to reduce the discomfort associated with having seemingly inconsistent beliefs about a person (2012, 116)

Interestingly, behavioral explanations appeal explicitly to the actor's beliefs and desires most often in two sorts of cases: one, "when they are motivated to portray the behavior in a positive light" (Andrews, 111, citing Malle 2004; Malle et al. 2007), and two, when they are explaining their *own* behavior (Andrews 111, citing Malle 2004; Malle et al. 2007). Assuming that we are regularly motivated to portray ourselves in a positive light, there seems to be a unity behind these cases: attributions of propositional attitudes are added to explanations of human behavior in order to portray it positively. Andrews suggests that,

³⁵ Andrews, K. (2012). *Do apes read minds?: Toward a new folk psychology*. MIT Press.

A better account of the function of mindreading is the development of sophisticated moral abilities such as the justification of behavior, which helps to explain both how a theory of mind facilitates group living and how it facilitates the development of technological advances” (2012, 218).

The function of evolving a theory of mind, she argues, is offering *reasons*, not predictions, particularly in anomalous cases where some norms are violated—so belief attributions presuppose a background system of norms, and are made to explain norm violations. In practice, people do not attribute propositional attitudes in explanations of behavior unless the behavior has deviated from communal norms. The ability to explain behavior at all is developmentally prior to the ability to use propositional attitude attributions for explanatory purposes.

While observers more often explain behavior by appealing to the causal history of the individual, and actors more often explain their own behavior by appeal to beliefs and desires, there are some conditions in which observers do explain others’ behavior via belief/desire attribution (Malle, 2004; Malle et al, 2007). Malle has found that observers will attribute beliefs and desires to actors when they are motivated to portray the behavior in a positive light. I suggest that when a group member acts in an anomalous manner, others in that group will be motivated to portray that behavior in a positive light, and hence the explanations that are generated will tend to be reason explanations in terms of propositional attitudes³⁶. (p. 438)

Therefore, it’s implausible that the ability to attribute propositional attitudes arose uniquely to explain and predict behavior. Instead, Andrews argues that the fact that propositional attitudes are invoked primarily in anomalous cases of norm violations points to a normative, rather than descriptive, function of mentalistic

³⁶ Andrews, K. (2009). Understanding norms without a theory of mind. *Inquiry*, 52(5), 433-448.

vocabulary. Tanney also focuses on how attributions of beliefs and desires are most at home when making sense of erratic or atypical behavior. In the article *Reason-Explanation and the Contents of the Mind*,³⁷ Tanney has us reflect on times we are puzzled by something another person does. Often more information about the circumstances is what demystifies what we've witnessed – why did the woman run frightened from the building? --- Because the building was on fire. In many cases, we are satisfied with the explanation without having to know anything about the person's mental states at the time – such as she *believed* the building was on fire and *desired* not to burn. Such attributions would be unnecessary to make sense of what we saw. Yet, there's a longstanding tradition in philosophy to insist that belief-desire pairs are necessary for appropriate explanation of what occurred, that something hidden with the agent must be uncovered to learn the true explanation of the behavior. But what if the building *hadn't* been on fire, and we saw the woman run screaming out the door. We might conclude that she *believed* that the building was on fire. This would also demystify the initially puzzling behavior. This affords us the psychological satisfaction of having made sense of the situation, while also painting the woman's behavior in a reasonable or positive light – she did what anyone in her position would have done. If the building had been on fire, we would not have had to appeal to propositional attitudes, but since it was not being able to do so afforded the opportunity for justification. As Tanney points out,

It does not follow from the fact that some explanations mention the agent's conception of the circumstances that all explanations do – even, say, when there is no misconception involved...I have

³⁷ Tanney, J. (2013). *Rules, reason, and self-knowledge*. Harvard University Press.

suggested that (in the normal case) an appeal to the circumstances is often enough. Now let it be conceded that sometimes we do have to appeal to the agent's conception of the circumstances in order to discover why she acted. But I shall deny that in appealing to how she conceives her situation, we must be homing on something inner. (p. 137)

Tanney denies that there must necessarily be something 'inner' when we invoke propositional attitudes in explanations of behavior because she denies that the language of propositional attitudes aims to *refer* to anything at all. Instead, invoking mental states in explanation plays a normative function within a sense-making pattern that we all begin learning from an early age.

In being able to communicate with us about her reasons for acting, such an agent shares a common social training in the activities or practices upon which our ability to see an event as an action rests. And this shared social training includes training in discourse about reasons --- a discourse that further enriches the whole character of concepts of reason, action, and agency by, among other things, manifesting the connections between these concepts and those of wanting, intending, believing, hoping, and so on. Someone trained in this discourse is able to make much finer-grained distinctions in talking about her reasons for acting than we may ever be able to do simply by watching her.

These shared social practices enrich our ability to rationalize and justify our behavior, and Tanney's insights are well supported by the empirical data on how explanations, justifications, and predictions of behavior are commonly made. But this prompts a new functional question – why is the ability to justify or rationalize anomalous behavior useful? Returning to Andrews psychologically informed analysis, we find that the ability for such justifications of violations of social norms to enable other group-members to understand and accept innovations (including technological innovations). One might also more broadly appeal to a reduction of

conflict among group members—for if someone’s action, that at first looked simply cheeky, offensive, or destructive, comes to be understood as reasonable given the agent’s beliefs and desires, rancor is reduced, and people are better able to live together. One can see this also in the typical difficulties faced by those on the autism spectrum, who often have trouble maintaining social relations and avoiding conflict—difficulties that go along with difficulties in making standard attributions of beliefs, desires, etc. to their neurotypical peers.

Since recent work in cognitive science has challenged the popular conception of propositional attitude attribution as primarily functioning to explain/predict behavior, we’re tasked with uncovering the role that such attribution *does* play in our social interactions. One suggestion is that mental discourse, particularly that involving propositional attitude attribution, allows us to locate human action within a web of reason that speaks to the rationality and reliability of our character. Social life is complicated in the contemporary world, where we can come in contact with more people in a single day than someone living before the industrial revolution would have seen in his or her whole life. But even in times and places where one’s social sphere is constituted by a relatively small community, group living presents special kinds of challenges. Any deviation from ‘normal’ ways of doing things can leave an individual open to scrutiny and reprimand, whether it is moral norms like honesty and loyalty, or matters of etiquette like eating with a fork instead of one’s hands or removing shoes before entering someone’s home. Having access to propositional attitude concepts affords unique abilities to make sense of norm violations in ways that forestalls censure, which is crucial to protect

one's reputation within the group. In all group living, a positive reputation creates more trust, and therefore more opportunity and potential for social success. When you are able to show that your behavior makes sense, given your history and how the world appears to you from your perspective, you signal to others that you are reliable and rational. A so-called 'theory of mind,' more specifically conceptual capacities related to propositional attitude attribution, allow for such nuanced signaling. In the absence of appeal to your beliefs, desires, intentions, hopes, and doubts it would be extremely difficult, if not impossible, to communicate to others in an effective way how your behavior makes sense from your perspective³⁸.

An approach to understanding propositional attitude attribution in this more normative, as opposed to descriptive, way can be found in the work of Daniel Hutto and Julia Tanney. Hutto's work is in agreement with Andrews' normative approach towards propositional attitude attribution. He too denies that appeal to a person's beliefs and desires plays the quasi-scientific role of explaining and predicting behavior. Hutto says,

...many so-called friends of folk psychology have overstated and misunderstood its role in social cognition and our lives more generally. First, they typically see it as more basic and far more pervasive than it is. We have many other more basic means of conducting social interactions. These yield neither predictions nor explanations *per se* but instead involve recognition-response patterns that generate 'embodied expectations'. In 'normal' contexts these are not only quicker but also far more powerful and reliable ways of relating to others and navigating social dynamics...Many of our routine encounters with others take place in situations in which the social roles and rules are well established, so much so that unless we behave in a deviant manner we typically have no need to understand one another by

³⁸ One's 'perspective' here is not indicating anything necessarily perspectival in a phenomenal or mental sense, but instead indicates the information at your disposal, your history, and other environmental factors

means of the belief/desire schema. More often than not, we neither predict nor seek to explain the actions of others in terms of their unique beliefs and desires at all³⁹.

In light of considerations such as these, Hutto goes on to advance what he calls the *Narrative Practice Hypothesis* (NPH) an account of the genesis of propositional attitude attribution. The main claim of NPH is that there's a distinct sort of story telling that engenders our ability to become conversant in propositional attitude talk – which includes becoming competent in applying principles that govern the interaction of propositional attitudes and other psychological states (such as perception and emotion) and being able to calibrate our attributions of propositional attitudes to relevant differences in particular cases involving a person's personal history, character, and current circumstances (2007, p. 48). According to this view, through exposure to person narratives throughout the earliest years of our lives, we learn how to apply and evaluate mentalistic predicates in ways that make sense of behavior and speak to the reliability and rationality of both others and ourselves. A 'person narrative' is a story about the events of someone's life told in the language of propositional attitudes and emotions. Cinderella has grown *unhappy* with how her stepmother and stepsisters treat her, and *wants* a chance to go to the ball and escape the abuse. The boy who cried wolf *wants* to trick his neighbors for fun, and so keeps setting off the alarm that a wolf is near. Finally, a wolf does attack the flock, but the neighbors do not *believe* the boy's cries, and so all the sheep are eaten. Hutto is careful to point out that while exposure to person narratives is the typical way we learn how to become conversant in propositional

³⁹ Hutto, D. D. (2007). The narrative practice hypothesis: origins and applications of folk psychology. *Royal Institute of Philosophy Supplement*, 60, 43-68.

attitude discourse, it is not the only possible way. Research exists that suggests autistic individuals tend to not internalize these narratives in childhood, which serves as a social obstacle in their early lives, but later in life are able to memorize explicit rules for applying folk psychological concepts, enabling them to compensate for their lack of insight into reasons why others act⁴⁰. Much different than the storytelling practice that Hutto argues is the standard way into propositional attitude talk, this purely “logical” and “calculated” competence with folk psychological terms tells us something interesting about what such vocabulary *does for us*. Hutto tells us,

Those who acquire their folk psychology skills in this way remain quite awkward in their dealings with others; they never fully develop a capacity to make sense of actions in the easy and familiar way that most of us do. The phenomenological differences are also salient. For example, Temple Grandin, an autistic individual who has, by her own account, succeeded in fashioning rules for understanding others in this way still ‘describes herself as like an anthropologist on Mars’. Of course, such feelings of estrangement have deeper roots, but the point is that these persist even after autistic individuals learn to master false belief tasks. This suggests that they never quite achieve the kind of understanding of others that is the norm for most people. Extracting this schema and becoming familiar with the norms for its application through experience with a certain class of discursive narratives is the culminating, non-negotiable requirement for a basic mastery of our everyday folk psychology abilities. Engaging in the relevant kind of story-telling practice is the normal route through which this practical knowledge and

⁴⁰ D. Bowler, ‘“Theory of Mind” in Asperger’s Syndrome’, *Journal of Child Psychology and Psychiatry*, 33, 1992, 877–893; S. Gallagher, ‘Understanding Interpersonal Problems in Autism: Interaction Theory as an Alternative to Theory of Mind’, *Philosophy, Psychiatry, Psychology* 11, 2004, 199–217. R. Eisenmajer and M. Prior, ‘Cognitive Linguistics Correlates of ‘Theory of Mind’ Ability in Autistic Children’, *British Journal of Developmental Psychology* 9, 1991, 351–364; F. Happé, ‘The Role of Age and Verbal Ability in the Theory of Mind Task Performance of Subjects with Autism’, *Child Development*, 66, 1995, 843–855

the normal route through which this practical knowledge and understanding is procured.

The Intelligibility View does not need to be committed to how we typically acquire the skills of attributing third-person. Without concluding that Hutto's explanation of person narratives is the standard way that we become conversant in propositional attitude discourse, the insights discussed in this passage can merely provide support for the suggestion found in Andrews', that propositional attitude attribution is playing a normative role of smoothing out social interactions to facilitate cooperation and ameliorate tension. When there's a breakdown in seamless propositional attitude attribution, as in the case of autism, we find corresponding obstacles to social interaction, and a rise in interpersonal tensions.

One might worry about the use of 'narrative' in this context, since the term has certain *fictional* connotations that might cast an eliminativist shadow over Hutto's account. If all we're doing is telling stories when invoking propositional attitudes in our conversations about human behavior, then in some more literal sense isn't it safe to say that there aren't *really* beliefs and desires? This concern is reasonable, although ultimately not a serious objection. I won't speculate on Hutto's behalf, since his stated motivation in advancing NPH is to better understand how we acquire and apply folk psychological vocabulary ontogenetically, and questions of phylogeny (let alone metaphysics of mind) do not appear to be of central interest in his work. But taking seriously aspects of Hutto's work as I move forward towards advancing a propositional attitude nondescriptivism that has certain metaphysical implications, I want to stress that

'narrative' here should not be read as 'fictional.' Constructing a narrative has to do with connecting events in a meaningful way, and there's no necessary link between such an endeavor and building a *fictional* account of events. More importantly, we can drop the categorization of the view as 'narrative' altogether, and more carefully say that according to both Andrews and Hutto, the ability to attribute propositional attitudes plays the normative function of making behavior intelligible in a pro-social way that may arise through narrative generation, or through other means.

Coming at the topic from a different direction than that of philosophy of psychology and psychiatry, as in Andrews' and Hutto's work, Julia Tanney presents a collection of articles centered on the them of questioning popular assumptions in philosophy of mind in *Rules, Reasons, and Self-Knowledge*.⁴¹ In the course of this investigation Tanney presents a strong case for the often overlooked normative dimensions of mentalistic discourse, central among these the practice of making ourselves *intelligible*. Tanney understands propositional attitude ascriptions as importantly functioning in our explanations of behavior not in a causal sense, as claimed by descriptivism, but instead as a way of *making sense* of a person's actions in a hermeneutic, or interpretive, way. On this Tanney says,

The acknowledgement that an individual has beliefs and desires that cohere with her action, as described, is to confirm that we have pinpointed an overall sense-making pattern – a dimension along which her doings can be assessed. Doing this, however, does not require that we construe her beliefs and desires as

⁴¹ Tanney, J. (2013). *Rules, reason, and self-knowledge*. Harvard University Press.

independent existents that play a role in the production of behavior. (2013, p. 8)

But understanding attributions of propositional attitudes in this way is not to *eliminate* propositional attitudes, but instead to locate them in their rightful territory not in the land of nonphysical substances, as the Substance Dualist argues; nor the land of neurobiology, as many contemporary physicalists claim. Instead, Tanney places propositional attitudes, and the attribution thereof, in the social dimension of sophisticated creatures such as ourselves,

To attribute propositional attitudes, I suggest, is not to name a state of the agent the nature of which threatens to be mysterious unless we can construe it as emerging from her brain and the brain's causal relations to the environment. It is rather to put a marker down on a particular pattern of descriptions of her thoughts, actions, and sayings that puts them in an understandable context. The ability to see actions as fitting into familiar patterns comes about through training and shared forms of life: a kind of acculturation which enables us to see actions in new ways. (2013, p. 9)

Making behavior understandable through the language of propositional attitudes, we're enabled as individuals living in socially complex environments to live together better by sharing patterns of interpretation. Although spelt out in different ways, *intelligibility* is the logic of propositional attitude talk, according to Andrews, Hutto, and Tanney. And much philosophical and scientific confusion has arisen from the conflation of hermeneutic explanation and causal explanation. Understanding propositional attitude discourse in this nondescriptivist way, which enjoys by methodological, theoretical, and empirical support, offers a promising path forward. The Intelligibility View is *nondescriptive* since it does not take propositional attitude terms to be e-representations, reliably covarying with certain

events in the world and used to *track* those events. But there is a (deflated) sense in which propositional attitude terms gain a certain descriptive role, I'll come back to this in Chapter 5, where I argue for the metaphysical implications of the Intelligibility View.

Suppose then that we maintain that the fundamental function of propositional attitude attributions lies in *making human behavior intelligible*, I'll call this the Intelligibility View. One hope is that some of the valuable insights of historical nondescriptivism will be preserved, while their shortcomings are avoided. As with prior forms of nondescriptivism, it avoids the detractions of the direct observation account by not positing a distinctive faculty of introspection dedicated to observing one's mental states; when I say "I believe the tie is green" or "I didn't realize you were standing in line," my primary aim may well be to make sense of my behavior (selling you the blue tie when you said you wanted green, going to the cash register before you)—not to issue a report on the goings-on introspected in my mind. The Intelligibility View also avoids the difficulties the Indirect Observation mind-reading account, for it doesn't treat mental states as theoretic posits that might fail, and that are needed for quasi-scientific explanations. This means we can stave off the threat of eliminativism by locating propositional attitude discourse in its proper place. Perhaps most importantly, this approach offers the hope of getting a unified account of the functions of both first- and third-person attributions. Whether John says 'I thought the tie was green' or I say, "John thought the tie was green," we may both be engaged in a kind of sense-making of his having sold it to the customer who wanted a blue tie—a justification that (if taken

up) would reduce the rancor and group-splitting between seller and buyer. So unlike Ryle's inference-licensing, Sellars' commitment-withholding, and Wittgenstein's expressivism, it gives just as good an account of why we would make first-person attributions as third; indeed it is clear why we have particularly good reason to make first-person attributions, insofar as we wish to reduce rancor towards ourselves, make ourselves look better, and protect our relationships with others.

At this stage, two objections are likely to arise. One is: we don't always use propositional attitude vocabulary with the aim of making someone's actions intelligible. Sometimes in attributing a belief, we might aim to vilify, as when we give a self-interested explanation of a politician's change in position on gay marriage ("she's just saying that now because she wants your vote"). There are two lines of response here. First, even in such cases propositional attitude attributions are playing a normative role, locating an individual's actions within a narrative that makes sense of their behavior. Even when belief or desire attributions seem to be serving the anti-social role of insulting or defaming someone, the more global function of the attribution may still be playing a pro-social function of bonding us with similarly minded people, and giving a coherent (if not disparaging) story about the behaviors we've witnessed. But of greater importance for any nondescriptivist account, we must make a distinction between the function of a type of discourse within the greater discursive practices of a group – whatever would answer "why would it be useful for language users to have that type of discourse *at all*," and the individual *use* of a particular piece of vocabulary

by a person at a certain time. – “How is this person using this bit of discourse *right now*.” Michael Williams makes this point nicely in the following passage, where he distinguishes *use* from *usage*,

An important first step towards answering this question is to distinguish two notions of “use.” There is use as *function*: what a word may be *used to do*, or what having it in the language is *useful for*. But there is also use as *usage*: *how* a word is *to be* used, which is a matter of respecting assertional and inferential proprieties.

I will discuss this distinction more fully in the next section.

A second objection is that this account doesn’t really get us away from the descriptivist assumption. According to the Intelligibility View, propositional attitude attributions work something like justifications or rationalization of human behavior. But if we’re going to appeal to the mental states in order to make sense of our behavior, a critic might argue, the best sense that can be made of our behavior is what really happened, or what’s *true*. But to be true, attributions of propositional attitudes must be taken to *describe* the mental states they attribute, and to *correspond* to those states. But then we must end up taking propositional attitude discourse to aim at describing entities it may or may not correspond to, even if the reason it serves us well is that it enables us to smooth social interactions in the ways suggested above. While there’s something to this line of thought, the objection is a bit quick: insofar as I am advancing an evolutionary explanation of the function of propositional attitude talk, it does not matter whether the attributions are true or not—what matters from the point of view of securing the evolutionary advantages is that they are taken to be true, or acceptable, or something along

those lines—accepted at any rate as intelligible, so that the rancor that otherwise might have been created is avoided, or the technological innovation taken up, etc.

Ryle had something to say about this,

Avowing “I feel depressed” is doing one of the things, namely one of the conversational things, that depression is the mood to do. It is not a piece of scientific premise-providing, but a piece of conversational moping. That is why, if we are suspicious, we do not ask, “Fact or fiction?” or “True or false?,” “Reliable or unreliable?” but “Sincere or shammed?” The conversational avowal of moods requires not acumen, but openness. It comes from the heart, not from the head. It is not discovery, but voluntary non-concealment⁴².

Andrews seems to agree with Ryle in one sense, saying

While science has truth as its goal, FP [folk-psychological] explanation does not share this single-minded focus,” folk-psychological explanations fulfill “a number of pragmatic goals” that “are more central to typical behavior explanation than is the goal of truth... we have no veristic requirement for whether something counts as an FP explanation or not. (2012, 116)

But goes on to argue that we should not be too quick in dismissing the veristic requirement, it’s just that truth shouldn’t be taken as the *only* goal of folk psychological explanation, showing that when we look to how people actually go about explaining behavior truth-aptness is just one part of the way these concepts function. In support of this insight, Andrews offers three features of satisfactory folk psychological explanation that put limits on what can count as appropriate propositional attitude attribution. These three features are as follows,

1. FP explanations are constructed by individuals as a response to an affective tension, such as a state of curiosity, puzzlement, fear, disbelief, and so on, about a person or

⁴² Ryle, *The Concept of Mind* page 87

behavior. This affective tension drives explanation-seeking behavior.

2. FP explanations reduce cognitive dissonance and resolve the tension that drives the explanation-seeking behavior; generating an explanation promotes a feeling of satisfaction.
3. FP explanations are believed by the explanation seeker and are not believed to be incoherent given the individual's other beliefs, regardless of whether the belief is true or consistent with those beliefs. (2012, 120-121)

By employing Andrew's requirements, we're able to preserve an essential part of Ryle's insight – that mentalistic explanations are not primarily in the business of providing truth-makers, while placing *some* limits on folk psychological explanations in order to distinguish between appropriate and inappropriate propositional attitude attributions. The Intelligibility View is able to block the objection that it characterizes propositional attitude attribution as mere confabulation in the service of playing a rationalizing role. It is not like an explanation of witch-talk that would speak of the function such talk served in weeding out non-conforming members of a community, while acknowledging that all attributions of witchhood are false. On the contrary, it seems crucial to make sense of the internal practices of propositional attitude attribution to acknowledge a distinction between cases in which such talk goes right and cases in which it goes wrong (through lies or mistakes)—and also to acknowledge the familiar asymmetries that suggest that propositional attitude attribution is less subject to mistakes in first-person cases than third. Andrews' three requirements help us make this distinction in a neat and systematic way. In the following two chapters, I will continue to show how the Intelligibility View respects the idea that belief

attributions may be true or false—although it doesn't do so by treating them as describing corresponding mental states.

§3.5 Chapter Conclusion

Despite the fact that few contemporary theorists defend Substance Dualism, the way we've come to understand the subject matter of philosophy of mind is still largely framed by the epistemological and metaphysical commitments of that view. Even when one rejects the claims that knowledge of our mind is in principle incorrigible, or that mental states are essentially private, without also rejecting the corresponding semantic commitment that mentalistic discourse functions to *describe*, one will still be constrained by the Cartesian framing of the topic in a way that will trap her in age old problems. The only way out is down, digging beneath the foundation of familiar debates and routing out the descriptivist assumption about how mentalistic discourse functions. Ridding ourselves of the descriptivist assumption creates the opportunity to think about the mind and its puzzles in an entirely different way. The Direct and Indirect Observation views considered at the outset of this chapter accept the traditional framing of questions about the mind, and lead to the corresponding traditional issues. The mind/body problem, problem of mental causation/interaction, problem of other minds, accounting for a faculty of introspection, and introducing the specter of eliminativism are all problems that result from a descriptivist position on mental discourse.

For these reasons, it is methodologically advantageous to see where nondescriptivism about mental discourse leads, even if the position seems initially implausible. What I've called the Intelligibility View is a synthesis of existing historical and contemporary nondescriptive threads in philosophy of mind under one thesis – the claim that propositional attitude attribution plays the primarily *normative* function of facilitating cooperation, ameliorating tension, and expediting technological/social progress by rendering abnormal or atypical behavior intelligible. This view owes a great debt to Ryle, Sellars, and Wittgenstein, and just as much to Andrews, Hutto, Tanney, Bar-On, Byrne, Thomasson and Finklestein. By incorporating insights of these researchers into one unified view of the function of propositional attitude discourse, I will be able in the next chapter to show how the new and improved nondescriptivism I've argued for here is able to face objections once devastating to similar accounts, and provide a path forward towards reunderstanding the metaphysics of mind.

CHAPTER 4: THE EMBEDDING PROBLEM AND OTHER OBJECTIONS

Overview

The proposal I have offered claims that despite surface appearances, the language of propositional attitudes does not aim to refer to anything at all – not to immaterial, neural, or functional states and processes; nor to theoretical posits, or dispositions to behave. I have argued that propositional attitude discourse serves to make behavior intelligible, to locate human action within a web of reasons that allows us to coordinate and cooperate as the socially sophisticated creatures we are. In the last chapter I gave theoretical and empirical support for this hypothesis, by showing how the Intelligibility View is able to overcome objections faced by the other contending positions on how propositional attitude attribution functions, while providing an answer that is most consistent with findings in the cognitive and social sciences.

Of course, no philosophical view is without criticism. In this chapter, I will first show how by deploying recent advances in nondescriptivism the Intelligibility View is able to surmount the infamous embedding problem. I will spend the remainder of the chapter dealing with other objections to the Intelligibility View, including how it fares in light of certain studies of propositional attitudes in cognitive science, what the view has to say about propositional attitudes in nonhuman animals, and the potential for the Intelligibility View to be extended to other types of mentalistic discourse like talk involving perception and emotion.

§4.1 The Embedding Problem

Even if I've managed to make a new functional hypothesis about propositional attitude attribution plausible by avoiding the shortcomings of earlier nondescriptivist views, no nondescriptivist account in any area of philosophy has credibility until it's able to manage the notorious embedding problem.⁴³ The force of the objection is found in the claim that in identifying the *function* of a certain expression we have not given its *meaning*. For simple sentences can appear embedded in the context of negations, disjunctions, conditionals and the like. In such contexts, the term is not being used in the nondescriptive way ascribed by simple uses, yet it seems we must allow that the term retains the same meaning in these embedded contexts (to deny this is to undermine the apparent validity of simple *modus ponens* arguments). In the most familiar form, it serves as an objection to expressivist views in ethics: If one thought that "Lying is wrong" serves to express disapproval of lying, and took that to give the *meaning* of 'is wrong', then it looks like it couldn't have the same meaning when embedded in "If lying is wrong, then getting your little brother to lie for you is wrong", since that evidently does not express disapproval of lying. On the other hand, if we accept that it has a different meaning when embedded in this context, it seems that modus ponens arguments like the following are invalid on grounds of equivocation:
Moral:

1. If tormenting the cat is wrong, then getting your little brother to torment the cat is also wrong

⁴³ Often called the Frege/Geach Problem in the literature, having first been developed independently by Peter Geach (1965), drawing on Frege, and subsequently by John Searle (1969).

2. Tormenting the cat is wrong
3. Therefore, getting your little brother to torment the cat is wrong.

The same kind of objection can be reconstructed for nondescriptivist views wherever they appear. We might think of the function of saying, "John believed the tie was green" was to justify his behavior (selling it to you), but belief attribution sentences can also appear in embedded contexts:

Mental:

1. If John believed that the tie is green, then he wasn't trying to trick you
2. John believed that the tie is green
3. Therefore, he wasn't trying to trick you

Here, in premise 1, the language of propositional attitudes is not functioning to make behavior intelligible at all (or performing any of the related social functions). Yet again, we cannot treat these attributions as different in meaning from attributions like those in premise 2 without treating apparently valid modus ponens arguments as equivocating.

To solve the embedding problem, a nondescriptivist must avoid *identifying* the meaning of the relevant term with its function, while still maintaining that function plays a central role in the term's analysis⁴⁴. One way to handle this problem is to give an inferentialist account of the meaning of the word 'belief,' identifying the rules it follows (including entry and exit rules for the discourse, which may connect it with perception and action). These rules remain in place even in

⁴⁴ reference Warren (2015) here.

embedded contexts, giving us an account of the meaning of the term that remains constant. Michael Williams develops this approach in detail in his paper *Pragmatism, Minimalism, Expressivism* by suggesting that pragmatists offer an EMU: Explanation of Meaning in terms of Use (2010, 323)⁴⁵.

The first step, following the pragmatist line, is to identify the functional component: asking what function individual uses of the term typically serve, and why it would be useful to have such terms in our vocabulary at all. As stated earlier, it's important to make a distinction here between function of the term and individual uses. Williams suggests that we

...Can distinguish *expressive or performative function* from *utility*: the question of what we are doing in deploying the word in question (typically in a non-embedded declarative sentence) from that of the ends we can accomplish, or needs we are able to satisfy, by doing that. Making these distinctions generates a meta-theoretical analysis, according to which an EMU will involve four components, divided into two sub-groups.

A. Content-determining (CD) clauses:

(MI) Material-inferential proprieties (intra-linguistic role).

(EC) Epistemic character of the inferential commitments determined by (MI).

B. Functional (F) clauses:

(EPF) Expressive/performative (speech-act) function.

(U) Practical significance/utility of possessing a word with that function⁴⁶.

⁴⁵ Williams, M. (2010). Pragmatism, minimalism, expressivism. *International Journal of Philosophical Studies*, 18(3), 317-330.

⁴⁶ Williams, M. (2016) *Epistemic Justification*

The three components of the EMU are the material-inferential⁴⁷, epistemic, and functional clauses. The material-inferential clause specifies the inferential relations that exist between a given term and other terms within the language, as well as nonverbal relations the term may have to behavior. The epistemic component specifies when it is appropriate to apply the term at all, the rule that governs correct application of the given concept. The functional clause is split into two parts: the practical significance of having such a term in your language *at all* and the expressive/performative power that a given instance of the term's use might play. Beginning with the Functional (F) clause, the Intelligibility View specifies (U) by claiming that propositional attitude attributions serve to make behavior intelligible in ways that facilitate cooperation and reduce group tensions. Having such terms in our vocabulary is worthwhile because they enable us to coordinate complex social interactions that would be impossible without such resources. As for (EPF), I want to stress here a pluralism about particular speech acts that fall under the umbrella of propositional attitude attribution – a pluralism we can find emphasized in the work of the theorists I've invoked in this project – Ryle, Sellars, Wittgenstein, Andrews, and Tanney. In a given instance an attribution of a belief may be playing any of the following roles (this list is not meant to be exhaustive) - expressive, commitment-withholding, inference-licensing, explaining, predicting, comforting, warning, obfuscating, vilifying, rationalizing, etc. The plurality of propositional attitude attributions is well evidenced by the research and is impressive in its

⁴⁷ *Material* inferences are to be contrasted with *formal* inferences. The latter are a matter of syntactic rules of language, while the former do not rely on formal structure alone.

variety and scope. But we can see how all of the ways a person may be using a propositional attitude attribution at a time serve the more general utility of such discourse – all the speech acts listed here are ways to render behavior intelligible in socially significant ways. In this way the Intelligibility View improves upon earlier forms of nondescriptivism, what once looked like disparate analyses of the function of mentalistic talk are shown to actually share a common functional core.

Having articulated the Functional Clause of the EMU, the next step is to identify the rules of use governing the relevant discourse and thus determining the content of the terms. Following Williams (CD) includes an epistemological component (EC), giving the rules for introducing propositional attitude talk; and the material-inferential (intra-linguistic) component (MI), specifying the inferential relations between terms like belief, desire, and intention and the greater language of which they're a part. The two components of (CD) are the meaning-giving component, they tell us what the terms mean, even in embedded contexts in which they are not used to fulfill the standard function of the discourse. Finally, we highlight how the account is still functionally driven, by connecting the content determining clause of the explanation with the functional clause, making clear why a term with rules like these fulfills the stated function.

I will begin with an EMU for belief and go on to articulate explanations of meaning in terms of use for two other key propositional attitudes, desire and intention. This is not an exhaustive treatment, terms such as *consider*, *doubt*, *judge* and *deny* will have to wait for future research. But by treating belief/desire/intention-attribution as a test case I hope to make the broader project

more plausible and show a direction to explore in the future, not only with other propositional attitudes but on to different types of mental states altogether, such as emotions, perceptions, and consciousness. Propositional attitude attributions occur in both first and third person, and the rules governing the introduction for these terms, as well as the inferential relations of the terms, will differ slightly for each case, and so I will specify (EC) and (MI) clauses for each.

Alex Byrne's recent work provides some helpful insight for articulating an epistemic clause for introducing propositional attitude talk.⁴⁸ Along the lines of the Sellars/Husserl view, Byrne argues that by starting with claims about the world, we are entitled to make certain inferences about our own beliefs—an entitlement which remains in place even if the original (world-oriented) experience wasn't veridical. Consistently with transparency theorists, Byrne holds that the inferences that enable us to come to know about our own beliefs don't rely on a separate capacity for inner observation, but only on our having certain perceptual and rational capacities that enable us to follow logical rules that permit inferences from world-oriented claims to mind-oriented claims. Byrne's strategy seems to perfectly provide the epistemological component of the EMU for belief states, a modified version of Byrne's epistemic rule (BEL) for gaining knowledge of one's beliefs:

(BEL): For an individual speaker, moving from p , to *I believe that p* is always acceptable. This is a free move, and the conclusion remains secure even if p turns out to be false. (2010, 112)

BEL is useful since it shows how we can gain knowledge of our own mental states that, unlike Direct Observation views, respects the intuitions of transparency and

⁴⁸ Byrne's relevant papers: *Knowing What I Want* (2011), *Knowing That I am Thinking* (2011)

requires neither positing nor a special faculty of introspection in order to gain knowledge of beliefs. Instead, we just follow epistemic rules such as that given in BEL. As Byrne emphasizes, this rule is good (knowledge-conducive) because it is *self-verifying*. (2011, 4-10) Any time one follows the rule the conclusion will be true. If a speaker accepts that it is raining, then she is permitted to infer that she believes that it is raining. Moreover, unlike mind-reading views, this gives us a way of accounting for a certain (plausible) form of distinctive first-person access and privilege. E1 can only be followed from the first-person point of view, because, to make the inference, the speaker must herself accept p and on that basis, given her possession of the belief concept, conclude that she believes that p . This feature helps account for the asymmetry between first and third-person attributions of mental terms often discussed in the literature as 'privileged access'. It's true that a subject is more likely to acquire knowledge about her own mental states than knowledge about the mental states of others. But according to Byrne, this isn't due to some special acquaintance she has with her own mental states. It is explained simply by the fact that this basic rule for introducing talk about one's own mental states can be followed only from the first-person perspective. I see that it is raining, and thus can come to believe *that I believe that it is raining*. But I cannot access the world-oriented claims that another person is in a position to evaluate, so am not licensed to follow E1 in the third-person case.⁴⁹ (Byrne, 2011, 4-9)

⁴⁹ While this rule puts us in a position to gain knowledge of our beliefs, Thomasson (2008) suggests how an analogous account might be extended to account for our knowledge of moods, qualia, and sensory modes. Each might require a somewhat different epistemic clause as part of its EMU, but the current example should give at least an idea of how this might go.

Filling out the rest of the EMU for belief, the content determining clause runs as follows:

(EC, first-person) From p , *I* am entitled to infer *I believe that p*. This inference is always good; it is a free move, and the conclusion remains secure even if p turns out to be false.

When we make attributions of beliefs to other people, we are taking world-oriented claims about their observable behavior and circumstances and moving to mind-oriented claims that play a justificatory role (the 'making sense of' above should be understood in a hermeneutic, not a covering-law/explanatory sense). Given that, the epistemic clause for third-person attribution is as follows

(EC, third-person) If some agent S 's believing that P would make sense of S 's behavior in the circumstances, that gives defeasible entitlement to infer that S believes that P . But such attributions are fallible open to repudiation, and are overruled by conflicting claims arrived at via E1.

It is worth pointing out that while there are two clauses to the epistemic component, there is a connection and priority ordering between them. E1 is epistemically primary over claims introduced via other routes. While E1 does not give us a route to third-person attributions (that's why we need E2), it does provide a 'check' on them, and one way of distinguishing apt from the inapt attributions, without averting to correspondence. (It needn't be the only way: they may also be repudiated by other evidence, or by replacement with an attribution that makes better sense overall of the behavior in the circumstances.) This relation between E1 and E2 also enables us to capture something of the idea of first-person privilege, without going for infallibility about one's own beliefs (any that are arrived at via a route other than E1 are given no special credence). This connection preserves Ryle's insight that dispositional properties related to certain mental terms allow us to settle disputes

and otherwise evaluate the aptness of propositional attitude attributions. The two-pronged epistemic rule also can do the work we needed to answer the second objection above. We are not treating belief reports as confabulations that just have to work to ease social relations; we do apply standards of correctness to them.

That leaves only the material-inferential component of the analysis of belief-talk to be analyzed. This is a matter of identifying the constitutive interrelations between belief and other concepts—other mental state concepts, behavioral concepts, etc. Here Ryle's observations become particularly relevant: that it is constitutive at least in part of our mental state concepts that attributions of mental states give us defeasible entitlement to make certain inferences. But, in light of the central role justification plays here and of the difficulties for the mind-reading account, we need to preserve the difference between the sorts of inference that are licensed by covering-laws (which Ryle also treated as inference-tickets) and the inferences licensed by mental state attributions. Rather than thinking of the inferences licensed as inferences about what the subject is disposed to do, we do better to think of them as inferences about what else it would *make sense* for the subject to think or do: that saying "John believes the tie is green" defeasibly entitles one to infer that it would make sense for John to offer it if asked to produce a green tie, to verbally report the color as "green" if asked, to deny that it is red or yellow, etc.

(MI): 'S believes that *p*' gives defeasible entitlement to infer that it would make sense for S to utter judgments that P, to act as if P were the case, to accept claims that P and reject claims that not-P, etc.

Those who were attracted to the mind-reading theory might fear that the functional account here doesn't give enough credit to the role of mental state attributions in explaining and predicting behavior. But we can now see how the epistemic and material-inferential components together enable us to account for the common use of mental-state attributions, in a slightly new key. Third-person mental state attributions, following E2, are those that *make sense* of S's behavior, this is a kind of explanation, though in the hermeneutic/interpretive sense, rather than in the sense of citing a cause in a covering-law. Tanney and Hutto's work discussed in the previous chapter are particularly relevant here, as we come to see how propositional attitude attributions build up a certain narrative of our behavior, or consistency of our character, in something like an explanatory role. But the explanation is less about tracking causes and effects of behavior, and more about painting a sensible (if not positive) picture of a person's behavior. Following the material-inferential component, such attributions entitle us to make inferences about what it would *make sense for* the believer to do in various circumstances, again engaging in something like prediction. To summarize, I offer the following EMU for belief attribution:

A. Content-determining (CD) clauses: Material-inferential proprieties (intra-linguistic role)

(MI): 'S believes that *p*' gives defeasible entitlement to infer that it would make sense for S to utter judgments that P, to act as if P were the case, to accept claims that P and reject claims that not-P, etc.

Epistemic character of inferential commitments determined by (MI)

(EC, first-person) From p , I am entitled to infer I believe that p . This inference is always good; it is a free move, and the conclusion remains secure even if p turns out to be false.

(EC, third-person) If some agent S 's believing that P would make sense of S 's behavior in the circumstances, that gives defeasible entitlement to infer that S believes that P . But such attributions are fallible open to repudiation, and are overruled by conflicting claims arrived at via E1.

B. Functional (F) clauses: Expressive/performative (speech-act) function and utility of possessing a word with that function

(EPF) speech acts involving belief attribution may (non-exhaustively) play expressive, commitment-withholding, inference-licensing, explaining, predicting, comforting, warning, obfuscating, vilifying, or rationalizing normative nondescriptive roles

(U) Belief attributions serve the function of making behavior intelligible in ways that enable us to better live together

It remains then only to show how a term governed by the rules articulated in the epistemic and material-inferential clauses would be able to fulfill its stated function, therefore connecting the clauses in a meaningful way. The epistemic clause makes it clear that in making belief attributions we are aiming to say how

the world shows up to the agent, clarifying her point of view on the world. The material-inferential clause entitles us to make inferences about what other states and behaviors would make sense for one to whom the world shows up that way. If her behavior made sense, given the way the world shows up to her, we have to that extent made her behavior understandable, intelligible — if not justified in an absolute moral sense, at least in a way that removes the feeling of anomaly or diminishes the impulse to blame and punish by reconciling the observed behavior with group norms. She is seen to only have done what anyone to whom the world showed up in that way might sensibly do — which, even if it doesn't match higher standards we might hope to bind ourselves to, at least diminishes the impulse to single her out for disapproval or punishment.

By not merely giving a functional analysis of belief-talk, but rather a full EMU, I'm able to answer the classic embedding objection to the Intelligibility View. You might worry that it is not a satisfactory analysis of propositional attitude attribution since sometimes belief attributions are used to vilify, or in embedded contexts, in which case they are not serving the function of justifying actions in a way that is socially useful. Unlike nondescriptivist efforts from the past, I do not take the function of the discourse to be the same as the meaning. And by making a distinction between the function of a particular speech act, and the utility of the discourse in general, we're able to see that not every belief attribution must be directly serving the an intelligibility function for it to still be governed by the same rules of introduction and inference. Instead, the function explains why it would be advantageous to have a term in our language that follows the rules articulated in

the epistemic component and the material-inferential component—rules that enable the term to fulfill its function. The meaning, however, is identified with the rules (of the epistemic and material-inferential components), and remains in place even in embedded contexts, or any other contexts where the terms are not being used in the service of their characteristic function.

The belief EMU is offered as an example of how the Intelligibility View answers the embedding problem. Making use of Williams helpful framework also give us the ability to even out implausible asymmetries between first and third person propositional attitude attributions found in earlier nondescriptivist accounts. Providing a full explanation of meaning in terms of use for other propositional attitudes will require less work now, and follow the same general framework. Treating the propositional attitudes as a monolith is common in the philosophical and scientific literature, and the Intelligibility View does claim a certain unity in the overall purpose of their attribution. But the rules that govern the introduction and inferential relations for each propositional attitude state will vary, and so an EMU for each would have to be given in order to separate the *meaning* of the term from its *function* and *utility*. But even some philosophers who are on board with a transparent analysis of belief, as given by the EMU above, still worry that the world-orientedness of belief, which allows us to follow E1, may not necessarily be plausible for other propositional attitudes. As Finklestein points out,

[I]t is difficult to claim that the self-ascription of belief [à la Evans] provides a model of self-knowledge that can be used in order to understand our awareness of our own, say, desires because there seems to be no “outward-directed” question that bears the

kind of relation to “Do I want X?” that the question “Is it the case that p ?” bears to “Do I believe that p ?” (2003: 161)

But Byrne thinks that this is overly pessimistic, and that there is a way to recast claims about our desires as claims about the world. In his paper *Knowing What I Want*⁵⁰, Byrne has us consider the example of deciding where to go for dinner, noting that when figuring out what we *want* what we’ll focus on is the comparative merits of the available options. Sushi sounds most appetizing, but the closest place is across town. There’s an Indian spot across the street, but they’re a little over priced. And the Indian place always gets so crowded, whereas the Sushi bar has a spacious dining room. Now I just need to weigh the convenience of walking across the street from the inconvenience of high prices and being crowded...in this way I decide that I want Sushi tonight, after all. Even if deliberations aren’t this explicit or rote, my eventual decision will be the option that I have most reason to choose. And what I have most reason to choose doesn’t obviously seem to be a matter of states of my mind, but instead circumstantial matters about cost, convenience, and flavor. Giving an initial pass at they type of conceptual rule we follow in order to arrive at knowledge of what we desire, Byrne offers the following:

DES* If ϕ ing is the best option, believe that you want to ϕ

Byrne acknowledges that this rule isn’t as good as the rule for belief, discussed above, since it both under and over generates instances of gaining knowledge about what one desires. Meaning, if this were the rule you followed to arrive at

⁵⁰ Byrne, A. (2011). *Knowing what I want*.

knowledge of your own desires, it would both fail to account for cases where you genuinely do desire something, and predict judgments that you desired something that you do not. If both options appear to have equal reasons in favor of them, then DES* will under-generate – I can know what I want without judging it to be the best option. Conversely, if I truly dislike all options given to me but must make a choice, I may just opt for one over the other, but not *want* to do what I've chosen. Even when acknowledging these problem cases, Byrne argues that DES* can be amended slightly

DES If ϕ ing is a preferable option; believe that you want to ϕ

Replacing “best” with “preferable” here can solve both problems of over and under generation of knowledge. To quell worries of circularity, ‘preferable’ here does not mean ‘what is desired,’ but instead “attractive, useful, or necessary relative to other options.” Now counterexamples may still arise, for instance, when someone resists the move from something’s preferability to her desire for it, like in the case of listlessness – you acknowledge that ϕ ing is preferable, nonetheless you don’t want to ϕ . I can say a couple of things about this, but the rule is still good. Judging ϕ ing preferable always gives entitlement to move to desiring to ϕ . And perhaps you do not really judge ϕ ing to be preferable in the case of listlessness, you recognize the good-making features of ϕ ing, but better still is the option of staying on the couch and doing nothing. Either way, the fact that you don’t always move from the judgment of ϕ ’s preferability to your desiring ϕ ing doesn’t challenge the goodness of the rule. From this I offer the following EMU for desire:

A. Content-determining (CD) clauses: Material-inferential proprieties (intra-linguistic role)

(MI): 'S desires that p ' gives defeasible entitlement to infer that it would make sense for S to express longing that P, to act as if P were preferable to other options, to accept claims promoting P and reject claims that discredit P, etc.

Epistemic character of inferential commitments determined by (MI)

(EC, first-person) For an individual speaker, moving from " ϕ ing is a preferable option", to "I want to ϕ " is always acceptable.

(EC, third-person) - For anyone, if some agent S's wanting that P would make sense of S's behavior in the circumstances, that gives defeasible entitlement to say that S wants P. But such attributions are fallible and open to repudiation, and can be overruled by conflicting claims arrived at via E1.

B. Functional (F) clauses: Expressive/performative (speech-act) function and utility of possessing a word with that function

(EPF) speech acts involving desire attribution may (non-exhaustively) play expressive, inference-licensing, explaining, predicting, comforting, warning, obfuscating, vilifying, or rationalizing roles

(U) Desire attributions serve the function of making behavior intelligible in ways that enable us to better live together

Additional analyses of propositional attitudes will follow suit. We are licensed to make attributions of intention to ourselves by moving from claims about the likelihood of certain actions factoring into bringing our goals about, to what we intend to do. Similarly for third-person attributions of intention, if we have awareness of someone's goals, and the information available to them, then we are defeasibly licensed to attribute intentions to that person – if I know that you've talked about starting your own business for years, and find out you just received a large sum of money and purchased a local laundromat, I am entitled to attribute the intention to open a business to you.

Fixing the meaning of propositional attitude terms with the rules that govern their introduction and inferential connections provides us with a way to answer the embedding problem. The content determining clauses of the EMU remain fixed for the relevant terms, even when they appear in embedded contexts. When we consider the argument discussed earlier,

1. If John believed that the tie is green, then he wasn't trying to trick you
2. John believed that the tie is green
3. Therefore, he wasn't trying to trick you

We can understand 'believe' in this embedded context to mean something like this: to say that John believed that the tie is green is to say that from judging that the tie is green, John *was* entitled to infer that he believed the tie is green. Insofar as he sincerely made that judgment, he was not trying to trick you. And he *did* sincerely make that judgment; therefore, he wasn't trying to trick you. Even though propositional attitude attributions *typically* serve intelligibility functions given by my account, I can make sense of atypical uses of the relevant terms thanks to the content determining clauses of the EMU. Moreover, this account gives a unified *transparent* treatment to both first and third person attributions of belief, equally plausible in both cases. The rules that govern the use of propositional attitude terms start with claims that do not require reference to mental states at all, but instead only to world-oriented content. Through a bracketing move we can become licensed, by following these rules, to speak about beliefs, desires, intentions, and doubts, but doing so is not an inward pointing way of referencing hidden or theoretical entities, but instead an outward directed way of facilitating social interactions. In this sense, it's an improvement of earlier mentalistic nondescriptive views, which were perhaps plausible for either first or third person attribution, but not both. The central point in this view of how to fix the semantic content of propositional attitudes is that meaning, at least sometimes, fixed by the rules that determine the term's correct usage.

§4.2 Squaring the Intelligibility View with Cognitive Science

According to the functionally driven Intelligibility View, the common way of understanding propositional attitude talk, at least the common way philosophers and psychologists have come to understand the discourse, is wrong. If propositional attitude talk seems to work the same as perceptual object talk, it's no mistake; it's likely that the introduction of such discourse was *modeled* on familiar descriptive language. But we shouldn't let the surface similarity between the two types of discourse mislead us. The language of perceptual objects aims to track things in one's environment. The language of propositional attitudes aims to render human behavior intelligible in ways that facilitate group living. And even though I've cited some research that seems either consistent with or supportive of this view, one might still worry that if the Intelligibility View is at odds with some of the currently most popular views in cognitive psychology, it will be impossible to square the view with empirical evidence. In this section, I will address this concern by looking at two influential studies involving propositional attitudes and try to square the Intelligibility View with the research.

§3.2 a – false belief research

Scientists interested in social cognition developed a way to measure a person's capacity for attributing false beliefs to people, an act that demonstrates an understanding of another's mind. In one widely discussed version of the study, researchers introduce child subjects to two dolls, Sally and Anne. The dolls

commence to play out a small scene, where Sally hides a marble in a basket and then leaves the room. While Sally is gone, Anne takes the marble out of the basket and puts it in a nearby box. Sally returns to the scene and the children are asked, “Where will Sally look for her marble?”⁵¹

Answering this test correctly, i.e. responding that Sally will look in the basket for her marble, is taken to show that the subject meets some minimum requirement for a theory of mind – the ability to attribute a propositional attitude to another person that is not simply an extension of the subject’s own perspective. Baron-Cohen’s original study was to explore social cognition in children with autism. And the study became quite influential, no doubt in part because it found a marked difference between neuro-typical and autistic children’s capacity for the task: 3 of the 16 autistic children (20%) answered correctly, compared to 13 out of 15 neuro-typical children (85%). Generally children aged four and older are able to pass the false belief test. These studies have been widely used to study theory of mind and its development. An objection to the Intelligibility View might be raised here – doesn’t the ability to tacitly attribute beliefs, evidenced in these studies, show that belief attribution *is* used to predict and explain behavior, and not necessarily for any social facilitation function?

Putting aside the question of what such research shows about theory of mind in autistic individuals – a question that has received a lot of attention⁵², there’s

⁵¹ Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? *Cognition*, 21(1), 37-46.

⁵² Ruffman T, Garnham W, Rideout P (2001). "Social understanding in autism: Eye gaze as a measure of core insights". *Journal of Child Psychology and Psychiatry* 42 (8): 1083–1094, Tager-Flusberg H (2007). "Evaluating the theory-of-mind hypothesis of autism". *Current Directions in Psychological Science* 16 (6): 311–315.

reason to suspect that the Sally-Anne study, and similar tests, might not be tracking belief attribution *at all*. When watching the puppets play out the scene described above, the children may be deploying heuristics such as “people look for objects where they were left” or basing their guess about where Sally will look on their own past experience – where do you look for something you want? Usually where you left it. The results of the study do not necessitate the conclusion that any attribution of belief has actually been made. Certainly nothing that threatens the Intelligibility View account’s analysis of the function of propositional attitude attribution, since there’s room within the view for explanation and prediction (of a sort – it makes sense, heuristically, of Sally’s behavior which might otherwise look silly), particularly in anomalous cases – maybe like when someone takes your marble and hides it while you’re not looking!

§4.3 Propositional Attitudes and Nonhuman Animals

We’ve come a long way since Descartes on the subject of animal minds, who proclaimed, “After the error of those who deny God, there is none that leads weak minds further from the straight path of virtue than that of imagining that the souls of beasts are of the same nature as our own⁵³” (p. 46). In philosophy and psychology the field of animal minds is of rapidly growing interest, with research in areas as diverse as Boden on animal perception (1984), Clayton on memory and

⁵³ Descartes, R. 1637, Discourse on the method of rightly Conducting One’s Reason and Seeking the Truth in the Sciences, Philosophical Writings of Descartes, Vol. 1, tr. John Cottingham, Robert Stoothof and Dugald Murdoch, 1985.

planning (2003), Griffiths on animal emotion (1997), Lurz on animal consciousness (2006), Rowlands on animal morality (2009). Researchers seem more than happy to leave the straight path of virtue to explore the souls of beasts. Given the progress that has been made in understanding the evolution of human cognition and the parallels we find in our nonhuman Earthmates, the Intelligibility View should say something about nonhuman animals capacity for propositional attitudes.

First, I'd like to remind you that the Intelligibility View is a meta-linguistic analysis of propositional attitude discourse. Although I will argue that there are certain metaphysical lessons to be learned from understanding propositional attitude attribution nondescriptively, the view that I've argued for here is an account of how attributions of propositional attitudes function within a human society. Since (as far as we know) there aren't any nonhuman animals that explicitly attribute beliefs to themselves or one another, one might think that the Intelligibility View has nothing to say about such creatures at all. But that would be dismissive of a very important question. One could easily conclude that since the Intelligibility View offers a nondescriptive view about how propositional attitude language functions, claiming that the language of belief and desire does not aim to *describe* anything at all but instead functions to promote cooperation and ease group tensions, that we cannot rightly say that animals have beliefs and desires since they are unable to make ascriptions, and their group living is so different from our own. However, I want to insist that the Intelligibility View is entirely able to make sense of the intuition that nonhuman animals have beliefs and desires. There's also a mistake

at the heart of this objection, it's not the ability to make ascriptions but to properly have them made of one that is at issue in the question of whether Ns have beliefs and desires.

Let's look back at the rule that governs the application of the terms "belief" or "desire:"

Belief:

(EC, first-person) From p , I am entitled to infer *I believe that p*. This inference is always good; it is a free move, and the conclusion remains secure even if p turns out to be false.

(EC, third-person) If some agent S's believing that P would make sense of S's behavior in the circumstances, that gives defeasible entitlement to infer that S believes that P.

Desire:

(EC, first-person) For an individual speaker, moving from " ϕ ing is a preferable option", to "I want to ϕ " is always acceptable.

(EC, third-person) - For anyone, if some agent S's wanting that P would make sense of S's behavior in the circumstances, that gives defeasible entitlement to say that S wants P.

Attributing beliefs and desires to nonhuman animals makes perfect sense given the third-person rules, since doing so may best make sense of the observed behavior. I used to keep treats on the bottom shelf of the pantry, but changed the treat spot over a year ago. Still from time to time, my dog paws at the cupboard door and whines. Why would she do that? Because she *believes* that the treats are in there, and she *wants* them. The first-person route into the concept doesn't apply, since nonhuman animals do not speak our natural language or share the same conceptual framework. The animals might have the entitlement to make the inference, but lack the relevant concepts. This answer might seem to be dodging

the question – “Ok, we’re licensed to *attribute* propositional attitudes to nonhuman animals, “ a critic might say, “but do the animals actually *have* beliefs and desires, according to the Intelligibility View?” I’d like to respond to this concern in two ways. If this objection seems to carry weight it’s only because the descriptivist assumption is still so strong in the way we parse propositional attitude discourse. If the language does not aim to refer, as the Intelligibility View insists, then continuing to look for underlying states of the subject is confused. And confused in the very way that has gotten us endlessly tangled in philosophical puzzles for centuries. Yet, as contradictory as it might initially seem, I do think that the Intelligibility View can lead to a sort of realism about the existence of propositional attitudes, and answer affirmatively to the question, “but do the animals actually *have* beliefs and desires, according to the Intelligibility View?” I’ll return to this topic in Chapter 3. For now I hope it’s enough to say that according to the view on the table we’re entitled to attribute mental states to nonhuman animals in certain cases, and that the meta-linguistic nondescriptivist position of my view is also consistent with the claim that the animals genuinely *have* propositional attitudes, although the ontology of such shapes shows up in a new light once we understand the function of the relevant discourse.

§4.4 Moving Past Propositional Attitudes

One final objection to the Intelligibility View is that it isn’t extendable to other sorts of mental states, and if I’m unable to provide a unified nondescriptivist account of

mental discourse in general, then the view is of limited interest and importance⁵⁴. Perhaps undergirding this objection is the idea that unification is desirable in a theory, and overall a unified view is always preferable. If the Intelligibility View, or perhaps a closely related nondescriptivist project, can't make sense of emotion attributions or phenomenal state attributions, then maybe all the Intelligibility View has accomplished is a game of philosophical Whack-a-Mole, pushing down the traditional metaphysical and epistemological problems of philosophy of mind only to have them pop right back up in other areas. In this section, I'd like to gesture at ways in which the nondescriptivist project can be made plausible for other types of mental discourse.

Even if you find the nondescriptivist case for propositional attitudes plausible, you might worry that attributions of emotions vary in important ways. In the first-person case, there's *something that it's like* to be furious at a friend for lying, or melancholy when hearing a song that reminds you of a forgotten love. And it seems like the language we use to ascribe fury or sadness, at least in part, plays the descriptive function of reporting on our emotional experience. It just seems implausible to claim that emotion attribution is not essentially descriptive. The Cartesian view makes the most sense here – we know emotions from personal experience, privately and intimately. We attribute emotions to ourselves as descriptions, or reports, of these inner states. It is only in extrapolation that we attribute emotions to others – while I can't feel your fury, or cry your tears, I can hear what you say, see what you do, and notice certain physiological indicators of

⁵⁴ Thanks to Brit Brogaard and Azenet Lopez for bringing this to my attention.

emotional states; and in virtue of these observations I come to describe emotional states that *you* may be in.

Although the Cartesian view is tempting in its intuitiveness, we should resist the descriptivist impulse present in the preceding analysis. Ryle didn't call the official doctrine 'official' without reason, and we've brought in a shared practice of treating "the mind and its contents" as *things*, so explanations of mentalistic phenomena in these terms will continue to draw us in. The nondescriptivist must maintain that intuition is an unreliable guide here, and is best rejected in favor of a careful analysis of the relevant conceptual practices. As stated at the outset of this project, the problems that result from a tacit commitment to descriptivism in philosophy of mind are reason enough to get outside the framework of the classic puzzles and seek a new way of approaching the topic entirely. In *Phenomenal Consciousness and the Phenomenal World* (2008) and *In What Sense Is Phenomenology Transcendental*⁵⁵ Amie Thomasson provides a promising path towards articulating a nondescriptivism about the *what it's likeness* of emotional and other affective states. In these papers, Thomasson develops an approach to recasting claims about our own 'inner' experiences – perceptual, emotional, and mood experiences – as *bracketed* claims about the 'external' world. The term *bracketed* here derives from Husserl's notion that the 'bracketing' involved in phenomenological reduction could enable us to shift from world-talk to phenomenological talk. Thomasson says,

⁵⁵ Thomasson, A. L. (2008). Phenomenal consciousness and the phenomenal world. *The Monist*, 91(2), 191-214.' Thomasson, A. L. (2007). In What Sense Is Phenomenology Transcendental?. *The Southern Journal of Philosophy*, 45(S1), 85-92.

How is it that (from a first-person point of view) we can acquire knowledge of what it's like to feel a certain mood or emotion? The strategy for the one-level theorist attain, it seems, must be to acknowledge that even in these cases what we are directly aware of is features of the world-as-presented by our experience –not, say, some intrinsic quality of the experience of depression or elation— and that we can acquire knowledge of the mood or emotion somehow by transformations backwards from the ways in which the world shows up to us. (2008, p.204)

These 'backward transformations' allow us to move from complexes of world-oriented information – “my best friend lied to me,” “I'm clenching my fists, flaring my nostrils, and my heart has sped up” to “I'm angry at my best friend.” As Thomasson puts it,

So emotions overall are understood as a complex of certain (belief-like) world representations, sensory representations of changes in bodily state, and functional dispositions to act in certain ways (or, in the case of moods, to engage in certain “styles” of behavior).

An approach of this kind provides the necessary first step towards nondescriptivism about what are thought to be the difficult cases of emotion and mood. If we're able to render such attributions as *transparent*, in the sense that they are ultimately world-oriented, then we are able to resist the strongest descriptivist pull – that the subject matter of emotion and mood attributions are essentially private, internal, states of the subject. Once we've called that assumption into question, and shown how we might recast talk of our moods and emotions as bracketed claims about world-oriented affairs, we're then free to ask what functional role such bracketing might play for us. Cartesianism is not inevitable, despite how alluring its descriptivist spell. While it's beyond the scope of my current project, I think there's reason to believe that studies into the logic of

emotion/mood attributions, along with research from cognitive science, will highlight the expressive and normative functions of the discourse.

§4.5 Conclusion

I've tried to show here how the Intelligibility View articulated in chapter 3 is able to stand up to both classical objections faced by earlier forms of mental nondescriptivism and more contemporary concerns regarding its consistency with existing science and potential for expansion. The Frege/Geach embedding problem, historically the scourge of all nondescriptivist projects, can be answered by fixing the meaning of the relevant terms in the rules of use for their application and inferential relations, while maintaining the importance of function in the account by showing how just these rules can realize the stated functional role of the discourse. Treating belief and desire as test cases for propositional attitude talk, I defended a function-first nondescriptive approach to such attributions, by showing how they can maintain their meaning in embedded and nonstandard contexts. Not only is the Intelligibility View able to overcome this historical obstacle, it is also the most parsimonious treatment of propositional attitude attributions, consistent with relevant research in social and developmental psychology. The view is able to justify our practice of attributing beliefs and desires to certain nonhuman animals, without any necessary inquiry into the neural sophistication or souls of such creatures. Finally, there's good reason to think that the Intelligibility View can extend beyond propositional attitude discourse, to talk about our

emotions, moods, and experiences, in a way that unifies mentalistic discourse under the umbrella of nondescriptivism. Doing so will help drive home the epistemological and metaphysical upshot of view – that the classic problems of philosophy of mind are rightly thought of as problems in the *framing* of the subject matter instead of difficulties with mental phenomena.

While my project has been focused mainly on language and linguistic practices, there's a certain approach to ontology suggested by the nondescriptive strategy. In the next chapter, I will thoroughly defend the claim that I've maintained throughout this project – that the Intelligibility View is *not* eliminativist, and what a proper understanding of mentalistic discourse delivers is a simple realism about things like belief, desire, and intention.

CHAPTER 5: EASY ONTOLOGY AND THE MIND

Overview

A function first account of mental discourse provides an understanding of the logic of propositional attitude attributions instructive in how to untangle certain persistent knots in philosophy of mind. When we become aware at the meta-level of the distinctive function such mental discourse plays, we start to see exactly why certain kinds of questions are inapt and how they smuggle in descriptivist assumptions that corrupt the way we think about the mind. Yet, a certain line of questioning still remains, one that is not interested in the *language* of propositional attitudes, but instead demands an explanation of the status of the mental states themselves. In previous chapters, I've argued that the primary function of belief-talk (and talk of intention, desire, and the like) is to serve the justificatory role of making our behaviors intelligible to us in ways that build a coherent narrative of our characters, facilitating cooperation and minimizing discord. Nonetheless, this view can be easily reconciled with the fact that we *do* refer to beliefs, desires, and intentions and some might reasonably ask what the fundamental nature of such states *is*. The Intelligibility View is not an error theory about propositional attitude discourse, claiming that although we *do* come to describe our beliefs and desires, such talk is a sort of convenient mistake. Instead, the strategy for understanding the nature of propositional attitude discourse I've developed in earlier chapters lends itself to a certain metaphysical view, one that is decidedly *not* eliminativist, although I

imagine it will take some work to prove this. In what follows, I argue that the Intelligibility View leads to (but doesn't entail⁵⁶) a certain stance on the ontology of propositional attitudes, one that is *realist* in the only sense that matters.

One concern that arises for the Intelligibility View, articulated in chapter 3 above, is that attributing beliefs, desires, and intentions only render behavior intelligible if those attributions are *true*— and this *requires* some reliable correspondence between the things we say and the mental states ascribed. More broadly, some might worry that any view in the nondescriptivist family undermines the idea that the relevant claims can be true. A version of this objection was famously raised by Donald Davidson (1963), who opposed the Wittgensteinian view that reasons are not causes. Although this concern is understandable, I will argue that it is ultimately misguided. The account I've given for the function of mental discourse allows that some belief/desire/intention ascriptions may be true and others false, but does so outside traditional correspondence terms. We don't need to think that the terms were introduced to correspond with some entities observed or posited, or that the truth of propositional attitude ascriptions is determined by taking this correspondence route. Nonetheless, we can have conceptual/linguistic practices in place for distinguishing the truth or falsity, aptness or inaptness, and sincerity or disingenuousness of propositional attitude ascriptions.

⁵⁶ There is no necessary link between the functional account of the discourse given in earlier chapters and the easy ontology approach developed here. But given what the Intelligibility View says about the function of propositional attitude discourse we're able to avoid more 'heavyweight' metaphysical views (like Representationalism and Dispositionalism) that bring along with them problems of mental causation, placement, and self-knowledge/knowledge of other minds that do not arise for the Simple Realist. For this reason, while Simple Realism is not *entailed*, for methodological reasons it ought to be preferred.

The EMU given in the previous chapter is ontologically conservative, in the sense that individual uses of ‘believes’ (for example) can be characterized without reference to the relevant propositional attitude. As Williams says,

the virtue of use-theoretic approaches generally – is that they are ontologically conservative. Representationalist explanations of meaning tend to inherit the apparent ontological commitments of the vocabulary under review. Thus a representationalist approach to moral predicates will tend to commit us *ab initio* to moral properties, and hence to metaphysical worries about their character. By contrast, the only antecedent ontological commitments of use-theoretic approaches to meaning are to speakers, their utterances, and so on: that is, to things that all parties are bound to recognize anyway. (2010, p. 320)

By following rule EC1 (first-person) we work up to belief-talk by beginning from statements about the world, and by following rule EC2 (third-person) we move from talk of relevant behavior/circumstances to the beliefs of others. We can thus come to see how belief talk may be introduced and proceed without having to think of it as serving a function of tracking distinct entities first observed through a quasi-perceptual route. We also, through the epistemic clauses, become entitled to accept some mental state attributions (arrived at through EC1 or EC2), and reject others (that don’t make sense of the relevant behaviors, when there is evidence of dishonesty, or when they conflict with claims arrived at via E1). As long as we adhere to a deflationary theory of truth (as a neo-pragmatist naturally does), we can then distinguish true from false belief attributions without commitment to any correspondence theory. This marks a crucial difference between the Intelligibility View and descriptivist accounts: while descriptivists can accept that in many cases claims about our own mental states or those of others serve to justify/rationalize

behavior, for the descriptivist the EMU is not conservative: we must refer to beliefs in order use these attributions in making sense of human action. By contrast, the Intelligibility View says that in the service of justifying/rationalizing actions we may employ conservative rules like E1 and E2 and thereby come to speak of beliefs without thinking of this talk as involving anything like tracking an observed or posited state of the world.

Saying that the EMU is conservative, and that belief talk may be introduced without having to first identify or posit beliefs to which the terms refer, emphatically does not mean saying that there are no beliefs. We may, using the rules identified in the EMU given in Chapter 3, legitimately come to say, "I believe that Hillary will make a great president" or "Herman believes that Bernie should get the nomination." From there, we may make a hypostatization enabling us to move from saying "I believe that Hillary will make a great president," to "I *have the belief* that Hillary will make a great president," now introducing a noun term for a mental state. But although, on this view, we may become entitled to refer to mental states like beliefs (and other propositional attitudes), the functional difference between discourse about mental states and discourse used to track butterflies or posit bosons is respected.

Provided that an approach along these lines can be generalized, pointing to these functional differences, and differences in the ways the terms are introduced, gives hope that certain classic ontological problems in philosophy of mind will come to show up differently. For we can refer to mental states alright on this model, but it's not at all like coming to track a new kind of creature or particle—

instead, it goes via a hypostatization from a statement that has a fundamentally normative function and is introduced without relying on any prior ability to detect or refer to beliefs.

Traditional worries about how propositional attitudes relate to physical 'things' might then be seen out of place, since according to the Intelligibility View we no longer treat 'mental things' as encountered or posited entities. We are able to see now how different questions about the causal efficacy or metaphysical structure of propositional attitudes are from questions about how some physically trackable things like tables relate to other physically trackable things like molecules. The problem of mental causation likewise arises most vehemently when we think of our terms for mental states as if they were describing observed entities, or positing theoretical entities to serve in causal explanation, and want to know how those relate to the physical world. When mental state talk is introduced as having a foundationally normative function, questions about how these 'things' we come to refer to could figure in causal explanations show up as, based on a category mistake, making as much sense as Ryle's famous joke, "She came home either in a flood of tears or else in a sedan-chair⁵⁷." I will demonstrate in what follows how a metaphysical strategy of this kind can be made plausible by discussing Amie Thomasson's work on easy ontology.

§5.1: Easy Ontology and Propositional Attitudes

⁵⁷ (1949, p. 12)

The study of metaphysics has a tumultuous history, having gone through periods of unfettered command as well as long stretches of academic ostracism. Metaphysicians have had to defend themselves against a few serious hits to the legitimacy of their projects in the past couple decades, and one notable example of this is Thomasson's meta-ontological strategy for deflating certain classic metaphysical debates. She offers a sustained argument for what she calls *easy ontology* in her 2015 book *Ontology Made Easy*⁵⁸. I'll give a brief exposition of the main moves in this strategy to show how a nondescriptivist about propositional attitude discourse can still maintain a firm realism about the entities in question.

The first step towards easy ontology in any domain is adopting a neo-Carnapian/Fregean analysis of the existence predicate. This means accepting that instead of thinking of existence claims such as 'Ks exist' as a normal, first-order predicate like "Ks are red," the existence predicate is taken to be second-order, used to mark when the concept K is instantiated. Thomasson, wishing to avoid difficulties that may arise by introducing the notion of 'concepts,' states the second-order existence predicate in terms of a noun's *application* and *co-application* conditions. She provides the following equivalence schema

E: Ks exist iff the application conditions actually associated with 'K' are fulfilled (2014, p. 86)

Application conditions should be understood as the basic rules of use that constitute a given term's meaning. A speaker gains proficiency with a term when she has mastered the rules that govern its use – rules that dictate when the term

⁵⁸ Thomasson, A. L. (2015). *Ontology made easy*. Oxford University Press.

should be applied and when it shouldn't. For noun terms, when application conditions are fulfilled, a speaker is licensed to say 'Ks exist'⁵⁹. What it takes to fulfill the application conditions for a given term will vary among noun types. Thomasson makes a distinction between *derivative nouns* and *basic nominative terms*. This will be important when making sense of the application conditions for propositional attitude terms. Derivative nouns are noun terms introduced into language by making use of more basic nouns already in place – their application conditions can be derived from claims that don't make any reference to the entities in question. Thomasson categorizes many social and institutional nouns under the derivative label – teachers are just people in certain roles, marriages and corporations come about by filing certain paperwork and paying certain fees. On the application conditions of derivative nouns, Thomasson says

These conditions may be stated in the form of rules that enable us to move from talk that did not make use of the relevant noun term (or any synonymous or co-referring term – though it may make use of other noun terms) to talk that does – introducing use of the new noun...a term like 'marriage' may be introduced and partially governed by the rule 'If two suitable people A and B visit the justice of the peace and fill in the relevant paperwork and say the relevant vows, then A and B got married.' Combined again with more basic facts – say, that Jack and Jill visited the justice of the peace, filled in the paperwork, and said the vows – we are entitled to infer that Jack and Jill got married, and so that there is a marriage, thus easily resolving the question of whether marriages exist. (2015, p. 100-01)

Social and institutional terms are not the only examples of derivative nouns. Using similar transformations, we can arrive at the existence of entities such as events, properties, and states. Given what I've already said about the rules that govern the

⁵⁹ I won't defend the deflationary view of existence here, the reader can find an impressive detailing of the view's benefits and plausibility by reading Thomasson's full argument.

use of propositional attitude terms in earlier chapters, propositional attitudes fall under the derivative category – we gain entitlement to make claims referencing such entities by following the governing conceptual rules that begin with uncontroversial claims making no such reference. In contrast to derivative noun terms, Thomasson also distinguishes *basic nominative terms and concepts*. This category refers to basic sortal terms, those whose application conditions do not depend on the existence of other sorts of entities in order to be fulfilled.

Before demonstrating how an understanding of the application conditions for propositional attitudes can help us arrive at easy answers to questions of their ontology, let me take a moment to give an example of another realm of philosophy where this strategy has been developed. Thomasson motivates approaching the topic of modality first from a linguistic angle by noting the various problems we encounter when addressing the issue in the traditional way of metaphysics – and the longstanding stalemate among theorists working in that area. So, she asks how we might forge a new understanding and explanation of things like ‘necessity’ and ‘possibility’ if we begin from a different entry point, language. Succinctly, the view is as follows: the discourse of modality does not serve to track some features of the world, but instead functions in a *normative* way. Talk about what’s *necessary* provides language users a convenient way to “express constitutive semantic and conceptual rules in the object language.” (page 2) Building off an analogy with the rules of games, where modal terms play an overtly normative role, Thomasson demonstrates how such terms function similarly in normal discourse. Consider the following,

Without modal verbs, rules can be stated using imperatives (e.g., in checkers, 'Black player: move first'). But we can also state them in the simple indicative form ('The black player moves first'). That shift to the indicative form brings some advantages: it enables us to express conditionals ("If the black player moves first then the red player doesn't move first") that we can use in making explicit our ways of reasoning among rules (to "The red player doesn't move first").

But without modal terms being introduced into our linguistic toolkit, there are limits to what we can express here. An indicative statement of a rule might be mistaken for a description, and further we have no ability to distinguish permissions from mandatory regulative statements. These limitations warrant the introduction of modal terms, which bring several advantages to rule-stating. Moving from "the black player moves first" to "the black player *must* move first" makes explicit that the rule is regulating and not describing a move in the game. Furthermore, with "the black player *may* move first" we gain the ability to express permissions, something we could not do without the introduction of a modal term.

Moving away from games, according to the normativist, metaphysical modal terms function very much the same way in ordinary language. On this view, modal language is understood as a set of terms and concepts expressing rules, permissions, and obligations. And from this linguistic account of the way modal discourse functions, there are important epistemological and metaphysical implications. Epistemically, the view explains how we can come to know modal facts, an epistemic issue at the center of debates about modality. According to Thomasson, we can see how modal facts can come to be known through the exercise of our conceptual competence. We can dispense with possible world talk

being anything more than a helpful device that adds expressive power to the language, and demystify the realm of the modal. By simply mastering the rules of use for the relevant terms, a language user can gain the ability to explicitly convey what those rules are, thus we can also get trivial inferences to possible worlds - and this is what *knowing modal facts* amount to. The modal case taken up by Thomasson illustrates how the easy ontology approach looks in practice.

The easy ontology strategy has been taken up by philosophers towards a number of ontological questions: Thomasson on ordinary objects and modal facts, Stephen Schiffer on properties and propositions, and Bob Hale and Crispin Wright on numbers⁶⁰. For example, Thomasson argues that the question of the existence of ordinary objects, such as tables, can be straightforwardly answered by beginning with a claim upon which both realists and eliminativists can agree:

- Uncontroversial claim: There are particles arranged tablewise.

We can move from there to introduce the noun term 'tablewise arrangement' as follows:

- Conceptual truth: If there are particles arranged tablewise, there is a tablewise arrangement of particles.

And yet for there to be a tablewise arrangement of particles seems to guarantee that the conditions are met for the ordinary application conditions of the term 'table' to be met. (Or we could, if there weren't one in our language, introduce the new noun term 'table' as short for 'tablewise arrangement of particles'). Thus, we can, by trivial inferences move to:

⁶⁰ Thomasson *Ordinary Objects*

- Derived claim: There is a tablewise arrangement of particles

And so,

- Ontological claim: There is a table⁶¹.

Although there have been metaphysicians who want to deny the existence of tables (Van Inwagen 1990, Merricks (2001) , citing some further criterion that needs to be met in order for an entity to exist, Thomasson argues that by understanding the rules of use that govern the application conditions of some concept or term, we can come to know exactly what it takes in order for an entity of that type to exist. And this mastery over the relevant concept tells us everything we need to know about the status and nature of the entity in question. In such cases, we can obtain positive answers to fiercely debated existence questions, and hopefully finally lay them to rest.

Now let's turn to the question of propositional attitudes. What would be the relevant conceptual transformation that allows us to move from claims involving no reference to a propositional attitude to a claim that does? In the case of belief, according to the Intelligibility Account given in Chapter 3, we begin with an uncontroversial world-oriented content, perhaps about a tie that a friend is wearing (perceptual claims are not the only world-oriented content from which we might begin. It could be a claim about the upcoming election or about the Battle of 1812.)

- Uncontroversial claim: The tie is green.

⁶¹ Ontology Made Easy (2015)

As discussed earlier, following EC1, we are *always* entitled to move from such claims to talk of what we believe. And so, we now introduce a claim with mind-oriented content:

- Conceptual truth: If you are in a position to affirm that the tie is green, you are entitled to affirm that you *believe* the tie is green

Making this first move, *ex hypothesi* is to introduce a term whose primary function is to play a certain normative role, and whose meaning is given by the epistemic and material-inferential clauses stated in the EMU. From this we get the following claim:

- Uncontroversial truth: I believe the tie is green.
- Conceptual truth: if a subject believes that p , then there is a belief that p

And finally, by engaging in a hypostatizing transformation (or a bit of nominalized paraphrasing) we arrive at the claim:

- Derived claim: I have the belief that the tie is green, therefore there are beliefs

These transformations demonstrate how we can move from language that originally contains no reference to beliefs, but arrive at such reference through a nominalization of the conceptually primitive *nondescriptive* belief-talk.

This might initially seem like a bit of philosophical sleight of hand, arriving at the existence of beliefs based on nothing but mastery over the relevant conceptual and linguistic rules. But understanding the nature of derivative nouns renders this intuition wrongheaded – the application conditions for a certain class

of noun terms may be stated by showing how to derive statements involving no reference to the entities in question to statements that *do* make such reference – that’s just the nature of such terms. If someone thought that there needs to be *more* to a marriage than two suitable people taking certain vows and signing certain paperwork, he would just be mistaken about what it takes for a marriage to exist. A nondescriptivist claims that a similar story must be given for propositional attitudes – if a person successfully follows the rules that govern the application of the terms, attributing a belief that-*p* to himself when he is in a position to accept that-*p*, or in the third-person case attributing a belief to another party based on the requisite verbal and behavioral evidence, then there *is* a belief. And anyone who thinks that some further requirement must be met in order for beliefs (and other propositional attitudes) to exist is simply mistaken.

An insistence that some further requirements must be met in order for beliefs to exist is an extremely prevalent view in philosophy of mind [Kim (1989), Fodor (1990), Dretske (1981), Shoemaker (2001)]. A critic might well argue that if propositional attitudes according to the Intelligibility Account are unable to play certain causal roles, then the application conditions for their existence have not been met. Or, as Kim says in his 1989 *The Myth of Nonreductive Materialism*,

...to be a mental realist, your mental properties [say] must be causal properties – properties in virtue of which an event enters into causal relations it would otherwise not have entered into. (1989, 279⁶²)

⁶² Kim, J. (1989, November). The myth of nonreductive materialism. In *Proceedings and Addresses of the American Philosophical Association* (Vol. 63, No. 3, pp. 31-47). American Philosophical Association.

In the next section, I will try to diffuse this objection and show that setting the bar for the existence of propositional attitudes at causal efficacy is mistaken. There is empirical evidence that convincingly suggests that propositional attitudes are *not* causally efficacious. But we need not despair at the threat of eliminativism such research has been taken to raise. The expectation of causal efficacy for propositional attitudes is a product of descriptivism, one we can eradicate along with the descriptivist assumption. Doing so does not lead to eliminativism, but instead to a simple realism about the ontology and nature of propositional attitudes that is socially and psychologically significant – showing that beliefs, desires, intentions and the like exist in the only way we should care about.

§5.2: A Realer Realism?

Returning to Davidson's objection to interpretation-based accounts of reasons, we can now employ the lesson from the previous section towards a solution. Consider what Davidson says here,

How about the other claim: that justifying is a kind of explaining, so that the ordinary notion of cause need not be brought in? Here it is necessary to decide what is being included under justification. Perhaps it means only...that the agent has certain beliefs and attitudes in the light of which the action is reasonable. But then something essential has certainly been left out, for a person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it. Central to the relation between a reason and an action it explains is the idea that the agent performed the action because he had the reason. (1963, p. 691)

For Davidson, a primary reason consists of a belief-desire pair, so when he claims that *reasons* must be understood causally he is claiming that *propositional attitudes* must be understood causally. Examples handled by Davidson aim to support this point, as when he discusses Melden's case of the driver who raises his arm⁶³. There may be two perfectly good rationalizing explanations of why the driver does what he does, for instance the driver may have been signaling a turn or he may have been waving to a friend on the sidewalk. The only way to differentiate the *true* reason, according to Davidson, is to identify the cause of the action. It's clear from what Davidson goes on to say that the type of cause he has in mind is a mental event,

Of course there was a mental event; at some point the driver noticed (or thought he noticed) his turn coming up, and that is the moment he signaled...To dignify a driver's awareness that his turn has come by calling it an experience, or even a feeling, is no doubt exaggerated, but whether it deserves a name or not, it had better be the reason why he raises his arm. (1963, p. 12)

Understanding 'reason' here as a belief-desire pair (the belief that his turn is approaching and the desire to signal), we can see that on Davidson's view propositional attitude terms function to refer to mental events. But as I argued in Chapters 2 and 3, there's good reason to think that the relevant concepts *do not* function in this way. For one thing, it's not *necessary* to invoke propositional attitudes to explain (or predict) behavior, since such goals can ordinarily be achieved by appealing to the environment, history, and other non-mental elements of the situation. Why did the driver raise his hand? Because his turn was

⁶³ Melden (1961) *Free Action* London: Routledge & Kegan Paul

approaching (or alternatively, because his friend was on the sidewalk). As both Tanney and Andrews discuss at length, ordinarily we ascribe propositional attitudes in cases of breakdown or confusion, when the types of non-mental reasons we might typically invoke fail. And if it's unnecessary to understand reasons as causes, we should resist the urge to do so – since placing this causal constraint on propositional attitudes introduces problems that can otherwise be avoided.

One serious issue introduced by placing a causal constraint on propositional attitudes is our inability to distinguish the 'true' reason from the 'merely rationalizing' ones. This is not a knock-down argument against the causal constraint, but it does call into question the explanatory clarity causality was intended to provide. According to Davidson, there can be any number of rationalizing reasons, but only *one* of those can be the true reason, i.e. the cause. But what evidence could we have to differentiate the causal reason from the rationalizing reasons? We might ask the driver why he had raised his arm, and ordinarily an agent's avowal does carry significant weight. But we also must admit that the driver might be lying, confused, or amnesic, and that first-person testimony is merely *one piece* of evidence we have for accepting or rejected an explanation of behavior. And according to the Intelligibility View, even when the driver himself provides reasons in terms of beliefs and desires, what he is doing is making his behavior intelligible in a way that will satisfy inquirers. What would we do if we were trying to distinguish the true reason for the driver's action from the merely rationalizing explanations? Presumably we would add further detail – if not the

driver's testimony, then perhaps the fact that his friend was seen standing on the corner, or some brain scan of what occurred in his head prior to raising his hand. Yet none of this could uncontroversibly settle the matter, all the introduction of further evidence does is add more justificatory machinery. One way of understanding how we distinguish the 'best' explanation is the one that makes the *most sense* of the given situation. There's no reason to conclude that the 'best' explanation is the one that causes the behavior, just as there's no reason to think that what anyone is doing when giving reasons (whether onlooker or driver) is attempting to identify a direct cause. It seems as if Davidson and I are at an impasse when it comes to the question of which explanatory model we should place on propositional attitude attribution. This point is brought into focus by recent research in the cognitive sciences that show that people very often don't understand, or have access to, the causes of their behavior (more on this below). If we place a causal constraint on propositional attitudes, claiming that beliefs, desires, and intentions only explain behavior insofar as they identify the causes of behavior, then we would have to conclude that quite often people do not their own beliefs, desires, and intentions nor the attitudes of others. According to the Intelligibility View, this outcome is absurd! As long as the attribution of a propositional attitude plays the relevant non-descriptivist role, then using the Simple Realist strategy outlined above we can come to correctly say that *there is* such a belief/desire/intention. Placing the causal constraint demanded by Davidson onto propositional attitudes sets the bar far too high for the existence of such states, and given empirical data on the topic might lead us into eliminativism

(if it turns out propositional attitudes do not, in fact, play a causal role in the production of behavior). Of the two accounts open to us, on the one hand a Davidsonian picture that construes reasons as causes, and on the other a rationalizing account of reasons, both provide a consistent story of the type of explanation we should expect from propositional attitudes. But the Davidsonian picture leads directly into the problem of mental causation and the threat of eliminativism, while the hermeneutic-explanation model avoids both of these serious worries. For that reason, we can dismiss Davidson's objection to the rationalizing role of propositional attitudes, noting that his alternative does not really make progress towards the issues he sets out to address, despite the lasting influence of his initial arguments.

While I don't find Davidson's insistence on a causal constraint for propositional attitudes compelling, it nonetheless has been a mainstay for realist accounts of the mental over the last half-century. Realism in its most generic formulation is an affirmation that some entities whose existence has been challenged actually exist; a realist about Ks says Ks exist. When it comes to realism about mental states, Drew Khlentzos says that,

A realist about mental states or conscious experiences is one who holds that our world contains creatures who are sometimes in states of believing, desiring, remembering, perceiving, etc. and that the existence and nature of such states in no way contingent on our recognition of them. The world is as it is independently of what we think about it⁶⁴.

⁶⁴ Khlentzos, Drew, "Challenges to Metaphysical Realism", The Stanford Encyclopedia of Philosophy (Spring 2011 Edition), Edward N. Zalta (ed.)

There are a number of ways to be a realist about mental states, but a few basic conditions seem to enjoy agreement. Most notably, realists about mental states (Fodor 1985, Kim 1995: 202; 1998: 79, 90; Shoemaker, 1980) agree that mental states only exist if they are causally connected to human behavior – able to affect, and be affected by, movements of the body. This causal constraint on what it takes for there to be a propositional attitude, and what a satisfactory metaphysical explanation must look like, supplies us with a requirement for what it takes for propositional attitudes to exist in addition to the application conditions offered by the Simple Realist view articulated above. Referring back to the discussion of application conditions in the previous section, we can pose the problem of defining mental realism in this way – what are the application conditions for the term *belief* (for instance) that the realist claims must be met in order to correctly say that beliefs exist? According to the Intelligibility Account, we can infer that beliefs exist through transformations that move us from uncontroversial claims that do not mention beliefs, to claims that do. The uncontroversial claims we begin with are arrived at any time a subject follows EC1 in the first-person case, or a subject follows EC2 for the third-person case. Not only do we arrive at the existence of beliefs using easy ontology's approach, the function first analysis of propositional attitude discourse undergirding this metaphysical position addresses why we might have initially thought more would be required – descriptivist assumptions have had a strangle hold on the way we're able to conceive of the subject matter of the mind. The nondescriptivist can, and should, avail herself of the realist label. Beliefs (and other propositional attitudes) exist, and play important roles in social reality and

our self-conception. Nonetheless, some will still see this account as *lightweight* and demand that there is a *realer* realism whose requirements remain unmet by the nondescriptive simple realist solution. So what might the application conditions for a realer realist account of propositional attitudes look like?

The tenets of the Received View, as articulated by Mathews (2007: 1), echo a sentiment expressed nearly twenty years earlier by both Jerry Fodor and Jaegwon Kim. In his 1987 book *Psychosemantics: The problem of meaning in the philosophy of mind*⁶⁵, Fodor says,

Mental processes are causal sequences of tokenings of mental representations. (1987, p.17)

If causal efficacy of the type described by Fodor and Kim is the requirement for a view to count as mental realism, then the Intelligibility Account has no right to call itself realist. Because, although we do ultimately become licensed to make reference to beliefs, desires, and the like, such entities' existence is derived via a hypostasizing transformation from claims that more fundamentally serve nondescriptive intelligibility functions. This being the case, the foundation of propositional attitude attribution is dissimilar from language about bats or bosons – language whose main function is to *describe* certain elements of the world, either in a directly observed or postulated sense. An expectation of causal efficacy is misplaced when you understand the role that propositional attitude attribution plays, as argued in Chapter 3 and 4. There are two reasons to push back on Kim's

⁶⁵ Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. The MIT Press.

suggestion: a theoretical reason, and an empirical one. As far as the theoretical pushback is concerned, throughout this project I have shown ways in which the descriptivist assumption constrains the possible moves available within any debate about the mind. The expectation of causal efficacy is one of these byproducts – it is a result of taking too seriously the analogy between the contents of the mind and the contents of our desks and shopping carts. When we begin by tacitly assuming language of propositional attitudes is saliently analogous to talk of perceptual object, when we wonder how the entities referenced factor into the causal swim. This is tied up with the thought that such terms are what Huw Price calls *e-representational*, with the function of tracking and causally co-varying with objects in the environment. As Price says, such terms follow an

...*environment-tracking* paradigm of representation, dependent on such notions as covariation and (what Field calls) indicator-relations – think of examples like the position of the needle in the fuel gauge and the level of fuel in the tank, the barometer reading and air pressure, and so on. In these cases, the crucial idea is that some feature of the representing system either does, or is (in some sense) ‘intended to,’ vary in parallel with some feature of the represented system. (Usually, but perhaps not always, the covariation in question has a causal basis.) (2008, 13)

In dispelling descriptivism we also dispel the need to account for the causal efficacy of propositional attitudes, and see it as a misguided way of thinking about beliefs, desires, doubts, and intentions once the fundamentally *normative* nature of the attribution of these states is grasped. Such states don’t explain behaviors in a causal sense, although, in part, they do explain behavior in a hermeneutic sense. But the Intelligibility account in no way denies ordinary intuitions that our mental states play an important role in how we understand ourselves and others – when

you *want* a beer you get up and walk to the fridge because you *believe* there is beer in there. The reason that attributing these states to yourself makes your behavior intelligible is because it's exactly what someone in your position would do. As argued in the previous chapter, the *causes* of behavior may be opaque to us, both in the first and third person cases. But in ordinary cases where no norms have been violated, and nothing odd has occurred, the attribution of beliefs, desires, doubts, and intentions aren't necessary to explain why certain behavior occurred. Why did you just go to the fridge? Because there's beer in there. If, unknown to you, your wife had drunk the last few beers, then you might say, "well I *believed* there was beer in there." Attributing the belief to yourself in this case makes sense of your behavior – but the state of belief is more correctly characterized as a rationalization for, than a cause of, your action.

Aside from worries about damage caused by the descriptivist assumption in debates about the mind, recent years have provided a steady stream of empirical evidence that threatens the alleged causal efficacy of propositional attitudes and other mental states. While the causal efficacy of mental states remains the dominant view in psychology, a growing number of researchers insist that evidence supports the claim that mental states have no efficacy (Wegner 2002, 2004; Holton 2004; Libet 1985, 2001, 2004). The acceptance of the causal efficacy criterion for mental states has led a number of theorists to voice skepticism about the existence of propositional attitudes, while it has led others straight into eliminativism (or provided support for already eliminative-leaning philosophers). Those who accept that *real* realism requires causal efficacy and want to avoid such

skepticism or eliminativism towards the existence of propositional attitudes are put in a tough spot by recent research. If we take seriously the empirical evidence that seems to threaten the causal efficacy of mental states, we are left with few options. Few options, that is, unless we can justifiably conceive of the application conditions for propositional attitudes as not including a causal requirement. But is this possible? In what follows, I discuss findings in cognitive neuroscience that challenge the purported causal efficacy of propositional attitudes.

§5.3: Threats to the Causal Efficacy of Propositional Attitudes

The claim that states like *belief*, *desire*, *intention*, and *doubt* are *causally efficacious* is an empirical one, and should be treated as such. Even though many advocates of the Received View claim that the only way to make sense of the way belief-desire attributions function in ordinary explanations of behavior is to view them as causally efficacious, certain strands of neuroscientific research from the past few decades put pressure on the claim. In this section, I'll discuss some recent findings that seem to challenge the view of propositional attitudes as causes of behaviors.

One compelling field of research concerns people who have had their corpus callosum severed (Gazzaniga 1983, 2000; Swanson 2000). The corpus callosum is a thick bundle of neural fibers connecting the left and right hemispheres of the brain, and may have been severed by head trauma, or more often intentionally as a surgical intervention that lessens the negative effects of severe

epileptic seizures. The function of the brain structure is to allow the two hemispheres to communicate with each other. When the corpus callosum is severed, there is no longer communication between the two hemispheres of the brain, resulting in profound lateralization. As is often the case with physiological study, pathology can be incredibly instructive in the processes that occur with things *go right*. By observing cognitive impairment caused directly by the severing of the corpus callosum, we are able to see how the functional brain uses lateralization in its ordinary processes. Split-brain research began in the mid 1950s, and over the course of the following decades, researchers devised a set of telling experiments (Meyers 1956; Sperry 1961; Bogen 1962; Meyers & Sperry 1958; Nebes 1973; Corballis & Sergent 1988) The general experimental procedure is as follow: the split-brain patient sits before a computer screen and asked to focus on a small piece of paper taped to the center of the screen. Images are then flashed for exactly 100 milliseconds to the right of the paper. Since, generally speaking, each side of the brain controls the opposite side of the body, the images are made accessible only to the left hemisphere of the brain by being placed to the right of the paper – the left hemisphere controls language functioning as well. Under these conditions, the split-brain patient is able to verbally report what he has seen on the screen, for instance a square or circle. The patient fixates on the paper at the center of the screen again. This time, an image is flashed to the left of center. This directs the image exclusively to the right side of the brain. The right hemisphere cannot produce speech, and due to the severing of the corpus callosum is unable to communicate what it has seen to the left side of the

brain in order to deliver a verbal report. When the patient is asked what image, he has seen he says “Nothing.” The image is flashed again, but this time the patient is asked to *point* at the image, not verbally report on it. When the image appears to the right of center, the patient raises his left hand and points to where the shape had been. When the image is flashed to the left of center, the patient raises his right hand and points to where it had been. This seems to confirm that each hemisphere is able to see the shape when it is flashed to the opposite visual field, and each hemisphere is able to use the hand it controlled to point and respond to questions. But only one side of the brain can *speak* about what it has seen. This research was revolutionary in the early sixties when findings were first made⁶⁶, which twenty years later would be awarded a Nobel Prize.

In the years that followed, additional patients were studied, and new discoveries about the lateralization of brain functioning were made. One famous case was that of Joe, who underwent the experiments described above and variations of these experiments. In one variation, simple line drawings are flashed either to the left or right of the center focus point. A car is flashed on the right, and when Joe verbally reports he has seen a car. A frying pan is flashed to the left of center, and Joe struggles, clearly frustrated, when he cannot report what he has seen. The researcher asks if he might try to draw the image that was flashed, and Joe promptly draws a crude frying pan with his left hand— the right brain is able to communicate pictorially. Next, a hammer is flashed to the right of center, while simultaneously a saw is flashed to the left. Joe is asked to draw what he has seen,

⁶⁶ Gazzaniga, M. S. (2014). The split-brain: Rooting consciousness in biology. *Proceedings of the National Academy of Sciences*, 111(51), 18093-18094.

and his left hand sketches the outline of a saw. The researcher asks what the picture is of, and Joe says “a saw.” “What did you see?” asks the researcher. “A hammer,” Joe reports. “Why did you draw a saw?” “I don’t know...I don’t know.” Joe says, clearly puzzled. Such studies help to establish the seeming independence and relative isolation of various cognitive functions. Yet another closely related set of experiments shed light particularly on the subject at hand – the causal efficacy of propositional attitudes. Gazzaniga explains the experiment,

In a related experiment, in the few patients with right- as well as left- hemisphere language, we lateralized written commands by presenting them tachistoscopically to the subject’s left visual field. In an example where the command was “laugh,” the patient laughed and, when asked why, replied, “I think it’s funny you guys come up and test us every month. What a way to make a living!” If the command “walk” was flashed to the right hemisphere, the patient would stand up from his or her chair and start to leave the testing van. When asked where they were going, the left brain might say, “I want a coke so I’m going into the house.” Again, the left hemisphere observes and interprets the actions of the isolated right hemisphere in order to create a verbal response⁶⁷. (2011, p. 119)

And another experiment involving a patient called VP,

Using a very elaborate optical computer system that detects the slightest movement of the eyes, we were able to project a movie exclusively to the left visual field. If the patient tried to cheat and move her eyes toward the movie image, the projector would automatically shut off. The movie her right hemisphere saw was about a vicious man pushing another man off a balcony and then throwing a fire bomb on top of him. It then showed other men trying to put out the fire. When VP was first tested on this problem, she could not access speech from her right hemisphere. When asked about what she had seen, she said, “I don’t really know what I saw. I think just a white flash.” I asked, “Were there people in it?” VP replied, “I don’t think so. Maybe just some trees, red

⁶⁷ Squire, L. R. (Ed.). (2011). *The history of neuroscience in autobiography* (Vol. 7). Oxford University Press.

trees like in the fall.” I asked, “Did it make you feel any emotion?” VP said, “Maybe I don’t like this room, or maybe it’s you, you’re making me nervous.” Then VP turned to one of the research assistants and said, “I know I like Dr. Gazzaniga, but right now I’m scared of him.” (2011, p. 120)

Due to neural plasticity, over time the split brain forges communication between hemispheres, and regains (an appearance of) unity among its functions. But researchers like Gazzaniga and Metzinger think the brain, when split, can tell us something important about what is going on in the neuro-typical case. The seamlessness with which patients are able to instantly generate a comprehensive narrative, not always – but sometimes – appealing to propositional attitudes or other mental states, must be present in the normally functioning brain. It’s unlikely that the brain gains this ability instantaneously once the corpus callosum is severed, it makes more sense that this *interpretive* function of the brain is typically occurring, but the split-brain patients afford the opportunity to control for lateralized presentation of stimuli in a way that lays the process bare. Gazzaniga concludes from his own research,

This experimental evidence merely illustrates a rather extreme case of a phenomenon that commonly occurs to all of us. Our mental systems set up a mood that alters the general physiology of the brain. In response, the verbal system notes the mood and attributes a cause to the feeling based on available evidence. Once this powerful mechanism is clearly demonstrated, given the complexity of real-life emotional stimuli, one cannot help but wonder how often we are victims of spurious emotional/cognitive correlations. (2011, p. 120)

But the correlations are only *spurious* if attributions of moods, thoughts, emotions, or propositional attitudes given in explanation of behavior purport to be *causal* explanation. If we understand the attributions to be playing the role that the

Intelligibility View claims they play, then they are not spurious at all. The individuals in the cases cited above are all seemingly attempting to render their behaviors intelligible, particularly when they are questioned by researchers. Presumably, they want to be helpful experimental subjects, and provide reasons for their observed behavior that make the most sense in the situation. The moods, emotions, and attitudes they attribute to themselves are not exactly spurious – they *do* make the behavior more intelligible. It's just that experimental conditions are different than everyday contexts. In the experimental conditions we know, in some sense, the *causal* explainers for the observed behavior – and therefore the subjects may appear to be “victims of spurious emotional/cognitive correlations.” But if we drop the requirement of causal efficacy for the existence of mental states, then the so-called cognitive correlations that are being drawn are not spurious at all. In ordinary situations they would play their role quite nicely – helping the patients make sense of their behavior in a way that wouldn't alarm or confuse the people around them, while also allowing the individual to maintain a self-conception of being a rational person. Although, if a person fails to follow the relevant rules for attributing propositional attitudes to himself, then the Intelligibility Account can agree that such attribution could be spurious, or at least fallible. Although some (Stitch 1983; P.M. Churchland 1986; P.S.Churchland 2002) have taken this research to be evidence for some sort of eliminativism, that conclusion is not inevitable. Instead, the Intelligibility View takes this neurological evidence as support for the idea that mental state attributions do not play a causal role, and shouldn't be expected to.

Realists about the mind (Kim et al) claim that mental states must be causally efficacious in order to exist. I have argued that this requirement is unfounded, its origin is the deeply rooted in the descriptivist assumption that talk of propositional attitudes either plays a tracking or positing role. When we drop that assumption, as the Intelligibility Account urges us to do, the plausibility of a causal criterion for propositional attitudes is reduced. But aside from academic debates, we do also have everyday ways of thinking of our beliefs, desires, doubts, and intentions as relevant to explaining our actions. The Intelligibility Account is perfectly capable of explaining such ordinary practices by showing how propositional attitude attribution enables us to make sense of behavior in socially conducive ways, and that such attributions can be more or less apt, depending on whether or not the appropriate rules were followed when making the attribution. Beliefs, desires, doubts, and intentions *exist*, and there are standards of correctness for their attribution detailed in their rules of use. The Intelligibility Account shows why the philosopher's expectation is erroneous, but makes good on ordinary practices of propositional attitude attribution. The empirical evidence provided in this section helps demonstrate that belief/experience reports sometimes can't have the causal role their subjects attribute to them—that they are making them up later, not reporting a cause, despite what they take themselves to be doing. This finding might lead a realist who demands causal efficacy to eliminativism, but that would be a mistake. Given the Intelligibility View, what we observe in split-brain patients makes perfect sense, they are attempting to attribute mental states in just the same ways, and for just the same sorts of reasons, that people generally do – to make

sense of confusing behavior. This needn't lead us to eliminativism if realists drop the criterion of causal efficacy for propositional attitudes.

§5.4: Chapter Conclusion

In this chapter, I have argued that the account of propositional attitude discourse given by the Intelligibility View leads to a Simple Realism towards propositional attitudes. Although descriptivism about mentalistic discourse has an undeniable pull, if we are able to resist the flawed way of interpreting propositional attitude discourse we're able to build an account of the metaphysics of propositional attitudes that has theoretical advantages over other prominent. The Simple Realism that stems from the Intelligibility View's nondescriptivism incorporates many of the benefits of competing theories, or explains why points that have been taken to be beneficial are actually misguided. Furthermore, Simple Realism avoids many of the objections of alternative accounts – its nondescriptive foundation sets it apart from other prominent views, particularly Representationalism about the mental. And finally, this easy ontological approach to propositional attitudes is arguably the most empirically adequate of all metaphysical analyses of such states. For it is able to make sense of evidence that challenges the alleged causal efficacy of propositional attitudes, while not taking such evidence to threaten the very existence of beliefs, desires, doubts, and intentions.

Chapter 6: Conclusion

i. Summary of Project

In the *Blue Book*⁶⁸ Wittgenstein remarks,

The man who is philosophically puzzled sees a law in the way a word is used, and, trying to apply this law consistently, comes up against paradoxical results. (1969, republished 1991, p. 27)

One cause of philosophical confusion comes from our expectations about how the relevant language works. Later, in the *Philosophical Investigations*, Wittgenstein says,

Our investigation is therefore a grammatical one. Such an investigation sheds light on our problems by clearing misunderstandings away. Misunderstandings concerning the use of words, caused, among other things, by certain analogies between the forms of expression of different regions of language. (PI 109)

I have argued that exactly this type of misunderstanding is the root cause of many difficulties found in contemporary philosophy of mind. By mistaking the everyday language about what we believe, desire, intend, and doubt to be similar in crucial ways to either perceptible-object-talk or scientific-positing-talk, we create the expectation that the 'objects' of propositional-attitude-talk must bear certain similarities to those other forms of discourse. When we realize that striking dissimilarities exist between mental states on the one hand, and perceptual objects

⁶⁸ Wittgenstein, L., & Docherty, P. (1991). *The Blue and Brown Books: Preliminary Studies for the 'Philosophical Investigation'*.

or theoretical posits on the other, we must admit that mental states cannot meet our expectations. And not only couldn't the sought after mental objects be similar to their analogs, but the way we have framed the investigation has ended up *multiplying* instead of solving problems. For instance, taking mentalistic discourse to be on par with perceptible-object-talk invites the following concern: perceptual objects are known through certain sensory modalities, so mental objects must be known through a distinctive faculty, too. But accounting for this faculty has been notoriously difficult. Also, perceptible objects have distinctive and publically accessible properties, and are governed by well-articulated laws – but mental entities cannot be clearly shown to follow suit. On the other hand, theoretical posits may fail to obtain, like phlogiston and caloric fluid, if better explainers or predictors for the phenomenon in question are found. Claiming that mentalistic language is fundamentally *theoretical* language introduces the threat of eliminativism, engendering the possibility that there are no beliefs, desires, intentions, or doubts at all. Eliminativists like to rest their argument on comparisons to things like phlogiston or caloric fluid, but we should be wary of the analogy here. While Eliminativists and cognitive scientists have given arguments for why we should regard mentalistic language as essentially theoretical, there are good reasons to reject the suggestion. No ordinary person, or society, has ever depended on the concepts of phlogiston or caloric fluid in the course of daily life, in the way that we all use mentalistic language to navigate our psychological and social realities. Although there are those who have made a career out of denying the existence of entities or phenomena whose existence is entirely uncontroversial to the world at

large (Van Inwagen 1990, Churchland 1985), there's reason to think that something has gone wrong when otherwise reasonable people arrive at such conclusions. The central claim of my project has been that taking propositional attitude language to be descriptive (in the tracking or positing sense) is what has gone wrong in debates about the mind.

Not all simple statements containing noun terms require that we posit objects in the world to explain what makes them true, in order for those statements to be true. Huw Price makes this point by having us imagine a child's puzzle book with stickers on the left page and a picture with many empty outlines on the right page⁶⁹. For each sticker – an umbrella, a javelina, the Empire State building – the child must find the corresponding outline in which to place it. Philosophy is a sophisticated version of this game. We have concepts/terms for certain things – moral principles, mathematical objects, modal facts, mental states – and find that they're *hard* to place in the natural world. Price calls this the *Placement Problem*,

The problem is that of 'placing' various kinds of truths in a natural world. We seem to have more truths than truthmakers – more stickers than places to put them. Since that puzzle thus turns on an apparent mismatch between the cardinality of two different sets, it should come as no surprise that there are three basic kinds of solution. One argues that the two sets can be matched, just as they are; that there is some non-obvious mapping that does the trick. The second argues that the problem arises because we have undercounted on the right, and that there are more truthmakers available than we thought. And the third argues that we have overcounted on the left, and that actually there are fewer statements in need of truthmakers than we thought (2011, p. 8).

⁶⁹ Price, H. (2011). *Naturalism without mirrors*. Oxford University Press.

As a nondescriptivist about propositional attitude talk, I, with Price, reject all three options. The first amounts to reductionism – a strategy whose main motivation in philosophy of mind seems to be in defending physicalism, which I argued in Chapter 5 is not, and should not be, a priority once we understand the *social* and *normative* nature of mentalistic talk. The second option takes too seriously the descriptivist assumption, and thus postulates that the natural world is more metaphysically abundant or complex than we currently have evidence to believe it is (without special pleading for the phenomenon one is trying to explain). And the third option seems to characterize eliminativist - to deny that the relevant statements about the mental should be taken as literally true. In the case of talk of the mental, I have argued that going eliminativist can only be the result of a misunderstanding of the nature of the discourse, or requiring mental states to have existence conditions far more demanding than is necessary.

Nondescriptivism has seen a resurgence in recent years, thanks to the work of people like Blackburn 1993, 2002; Price 2011; Williams 2010; Amie Thomasson 2015) who have carved out a way to give a function-first analysis of a given realm of discourse while epistemically grounding the semantic content of the term through the articulation of the rules of use that govern the term's application and inferential connections to other terms (and nonverbal behaviors). The rules of use *do* remain fixed throughout contexts, while the *use* of the term may change from its standard function. In this way, we honor the insights into the functions of various forms of discourse that nondescriptivism highlights (and in so doing preserve some of the lessons of historical predecessors), while making the view stronger and

capable of overcoming the embedding problem and other objections. In Chapter 3, I argued for the fundamentally nondescriptive function of propositional attitude discourse, that being able to speak about our beliefs, desires, intentions, and doubts affords us certain social benefits otherwise unavailable. These benefits include easing group tension when individuals transgress against community standards, allowing us to cooperate by 'seeing things' from another's perspective, facilitating innovation in technology and practice by easing the discomfort that arises when some new way of doing things or behaving is first introduced to the group, affording the opportunity to lessen our own responsibility by hedging commitment to reports we deliver, and in general making behavior intelligible in ways that creates a positive narrative about the rationality and reliability of our character so that we can better form relationships and protect our reputation.

There are theoretical benefits to seeing propositional attitude talk as fundamentally normative in this way. Dropping the analogy between mental state ascription and either perceptual-object talk or theoretical-positing talk prevents us from running head first into the problems that arise for such views. There is also a great deal of empirical evidence from cognitive science supporting the Intelligibility View's emphasis on social cognition as the correct level of analysis on which to explain mental phenomena. In Chapter 3, by employing the work of contemporary nondescriptivists (particularly Michael Williams) I demonstrated how the Intelligibility View is able to overcome the dreaded embedding problem, and tried to answer a host of other objections that might also arise for the view.

If all of this works, what we are left with is a meta-linguistic account of the function of propositional attitude talk as primarily *normative* and *nondescriptive*. The language of beliefs and desires was never meant to track little privately accessible things contained within our minds; nor was the language ever meant as pseudo-scientific theoretical positing language serving the function of helping us predict and explain behavior. The language of propositional attitudes as it exists in our everyday life is a bit of social lubrication necessary for sophisticated creatures like ourselves. But if that's true, then you might have a reasonable concern about the metaphysics of all this. Is the Intelligibility View a type of error-theory, after all? Telling us that we might *talk as if* we have beliefs and desires because it is convenient to do so, but when you really get down to it there *are no such things*?

I emphatically reject this construal of the project, and in Chapter 5 argue that the intuition that the Intelligibility View might be an error-theory is further evidence of the descriptivist assumption still at work in the way we think about the problems of philosophy of mind. To make this point, I show how existence conditions for the concept "belief" (or other propositional attitudes) can be found in the EMU provided in Chapter 3 – that the rules of use that govern the terms also provide all we need to know about what it takes for beliefs (and the like) to exist. Specifically, if we have attributed a belief to ourselves following EC1, or to a third party following EC2, then via a hypostatizing transformation, we can come to truly say that a belief exists. This *easy ontological* approach, adapted from Thomasson (2015), delivers a simple realism towards propositional attitudes. I then challenge the "realer realist" who charges the Intelligibility View as an error theory leading to

eliminativism to say *what further requirement* would have to be met in order for propositional attitudes to exist. The further existence condition cited by most “real realists” is causal efficacy; the critic will claim that since the Intelligibility View does not cast propositional attitudes as *things* (brain states, nonmaterial dualist entities, functional states, etc) then it is unable to give causal explanations of behavior and in practice amounts to elimination of what should be meant by *mental states*. I argue that a causal constraint on mental entities is both theoretically and empirically fraught, and we ought to reject the “realer realist’s” demand for causal efficacy to be among the existence conditions for propositional attitudes. If this argument is convincing, we are free to accept the Intelligibility View about the language of mental discourse, and the Simple Realism that fits so nicely with that meta-linguistic position. In doing so we exorcise a number of problems that have possessed philosophy of mind for so long, and free ourselves up to think about the realm of the mental in a very different way.

ii. Future Research

I have focused throughout this project on propositional attitude attributions and what we can say about the nature of beliefs and desires given a nondescriptivist understanding of language involving them. But the point of my critique of the presupposition of descriptivism that frames debates in philosophy of mind touches upon a number of issues that fall decidedly out of the scope of propositional attitude research. In this section, I will identify two topics that could

potentially benefit from the insight of nondescriptivism at the core of the Intelligibility View. The topics I'll discuss here are questions of consciousness and the self. Metaphysical and epistemological puzzles involving consciousness have strong ties and similarities to those involving propositional attitudes, and so the possible application of my argument into these areas seems like a natural extension as I work towards building a comprehensive approach to questions of the mind that go beyond our ability to attribute propositional attitudes. I will add that the work of dissolving or diminishing the problems of the mental for propositional attitudes might be thought of little value if we still need other (either more robustly realist or eliminativist) solutions to handle other types of mental phenomena. And so, I'd like to suggest here how nondescriptivism might be workable towards some other persistent problems in philosophy of mind.

1. *Consciousness*

Thought experiments aimed at challenging the adequacy of physicalism have dominated debates about the mind for decades. Jackson's Knowledge Argument (1986), Chalmers' zombies (1996), Nagel's *What is it Like to be a Bat?* (1974), inverted spectra (Shoemaker 1982; Block 1990), and Searle's Chinese Room (1980) have largely been taken to demonstrate the serious limitations of Behaviorism, Functionalism, and Identity Theory. The function-first strategy I developed towards propositional attitude attribution and related problems of metaphysics and epistemology will only be of value if it is able to weigh in on these

core issues of phenomenal consciousness. I have little more to say here than some vague gestures at how I believe the Intelligibility View may be extended into this area, but I acknowledge it as a necessary and logical next step for research.

A natural starting point for building a nondescriptive view of consciousness is to look to the function of the discourse, how do ordinary people *use* terms of consciousness like *experience*, *sensation*, and *appearance*? Establishing a foothold in linguistic/conceptual practice is needed in order to understand the rules that govern these terms. Of course, it might reasonably be argued that common usage of consciousness terms is not going to be relevant or helpful for making progress in scholarly debates. Brian Keeley traces the history of philosophical terms *quale* and *qualia* (2009), concluding that the terms have been deployed in messy and disparate ways from their inception in 1866 By C.S. Pierce to present day, quoting Tim Crane (2001) as saying⁷⁰,

To have a clear understanding of [the mind-body problem], we have to have a clear understanding of the notion of qualia. But despite the centrality of this notion [...], it seems to me that there is not a clear consensus about how the term 'qualia' should be understood, and to this extent the contemporary problem of consciousness is not well-posed. (170)

If Keeley and Crane are right, and the contemporary problem of consciousness is not well-posed, the nondescriptivist will say that we must begin by gaining a clearer understanding of the ordinary and academic concepts as they're used, and that this

⁷⁰ Keeley, B. (2009). The early history of the quale and its relation to the senses. *Routledge Companion to the Philosophy of Psychology*, Routledge Press, New York.; Crane, T. (2001). "The origins of qualia." *History of the mind-body problem*. T. Crane and S. Patterson. Eds., Routledge: 169-94.

will hopefully lead to untangling this knot of related puzzles. I plan to further develop the strategy deployed in the current project, excavate possible descriptivist assumptions framing these debates, aiming to show how they lead to inapt questions once the relevant language is properly understood.

Take Jackson's Knowledge Argument (1986), for example. Up until adulthood, Mary is kept in a black-and-white room, becoming the world's preeminent color vision scientist having never experienced color herself. This much is supposed to make sense to us. Mary finally leaves the room, and for the first time sees a bright red ripe tomato, and we're asked if she's learned anything new. If we concede that she *has* in fact learned something new, apparently there's a problem. Because Mary knew all the physical facts about color vision, so by definition she could not have learned any *physical* facts. She must have learned some *non-physical* facts, the argument concludes. Therefore, physicalism is false and phenomenal facts are essentially irreducible. For the descriptivist, red flags are going up everywhere in the framing of the Knowledge Argument. How might Mary have become conversant with color terms had she not encountered them in the standard way, i.e. by ostension? Perceptual phrases like "looks red" are descriptive in nature, predicates that attach to worldly objects like tomatoes, fire trucks, and poppies. Mary would have had no opportunity to apply the term red in the standard way, or learn the discriminative and inferential aspects belonging to that term. When you are a child, you point to a rose and say "look, red!" And your father might respond, "yes, that is red!" or "no, this one is *pink*." It's difficult at first to apply the term correctly, but through practice we master the occasions on which it is apt. Mary

has had no such experience. And so there's a certain conceptual capacity that Mary had no chance to exercise or hone, namely the capacity to apply the concept red on the basis of experience, although she has had the chance to apply that concept based on certain scientific knowledge. This leads me to think that David Lewis' famous Ability Hypothesis (1988) solution to the Knowledge Argument is on the right track. Whatever Mary learns when first seeing red is more likely know-how, as opposed to know-that. But there's another sense in which we should refuse to get into this argument at all, since the motivation behind the Knowledge Argument is either to disprove or defend physicalism. Not all mentalistic practices are reducible or meaningfully explained in terms of underlying physical facts. But that doesn't mean that the world must have some fundamental *non-physical* properties, it means that the level of analysis that is interesting and instructive for understanding the nature of mental phenomena is not at the level of functional or physiological facts. Maybe you can reduce all the choreography, music, costumes, and narrative of a ballet down to physiological and functional facts, but you won't have learned anything meaningful about *La Bayadere* in doing so.

2. *The Self*

Stepping outside of philosophy of mind proper, the Intelligibility View potentially has promising implications for the topic of the self. Striking similarities exist between debates over the nature/existence of the self, and debates about the nature/existence of propositional attitudes. First, there are certain structural

similarities that cause debates about the self to run parallel to debates about mental states. For instance, causal efficacy is often thought to be a requirement of the self, many try to reduce the concept of the self down to neuro-physiological structures/processes, and others have been driven to eliminativism/fictionalism about the self. The commonalities in structure suggest that the same sort of flaws in the framework of debates about the mind might be at work in debates about the self. Moreover, the two subjects are saliently connected. What we take our self to be is inextricably bound to psychological factors like personality traits, memories, values, beliefs, and temperament. If the Intelligibility View, or a closely related nondescriptivist project, can move past propositional attitudes on to other mental categories (a prospect I argued for in Chapter 5), then by extension what we conceive of as 'the self' might best be thought of as an abstraction over various nondescriptive normative processes.

Some philosophers, neuroscientists, and psychologists have taken strains of empirical research to have powerfully proven that the self is an illusion, that what we once took to be the center of our lives does not exist at all. Dennett appeals to split-brain research (discussed in Chapter 5) to show that, "A self is also an abstract object, a theorist's fiction"⁷¹. Thomas Metzinger agrees, claiming in his 2004 book *Being No One* that, that "no such things as selves exist in the world: Nobody ever was or had a self."⁷² Such skepticism towards the self is unfounded, and for one of the same reasons that Eliminativism misses the mark. The theorists

⁷¹ Dennett, D. C. (1992). The self as a center of narrative gravity. *Self and consciousness: Multiple perspectives*.

⁷² Metzinger, T. (2004). *Being no one: The self-model theory of subjectivity*. MIT Press.

quoted above seem to think that the existence conditions for “self” are unmet, and that science is once again disabusing us of our folksy illusions. But what exactly the criteria for selves *are* that allegedly have gone unmet are not requirements that we should burden the concept of ‘self’ with in the first place. Constraints that include causal efficacy, unified neural processing, and executive control over cognitive capacities are the requirements for a self to truly be said to exist, according to the skeptics. But if I consider ‘my self’ to be an abstraction – a nominalization or hypostatization over more foundational normative, nondescriptive practices – then there is no reason to deny the existence of such an entity. The details need to be worked out more fully, but just as in the case of propositional attitudes I want to stress that this sort of view is not a consolation prize, the realism we *have to* accept since the “realer realism” has turned out to fail. Instead, once the function of the relevant discourse and how concepts of mind hang together to create our social reality is appreciated for what it is, we can turn back to the question of what an individual self is with a better understanding of what is reasonable (and unreasonable) to expect.

BIBLIOGRAPHY

- Andrews, K. 2011. *Do Apes Read Minds?: Toward a New Folk Psychology*. MIT Press.
- , 2009. Understanding Norms Without a Theory of Mind. *Inquiry*, 52(5), 433-448.
- Armstrong, D.M. 2004. *Truth and Truthmakers*. Cambridge University Press.
- Bar-On, D. 2004. *Speaking My Mind: Expression and Self-Knowledge*. Oxford University Press.
- Baron-Cohen, S. Leslie, A. M., & Frith, U. 1985. Does the Autistic Child Have a “Theory of Mind”? *Cognition*, 21(1), 37-46.
- Bickle, J. 1992. Revisionary Physicalism. *Biology and Philosophy*, 7(4), 411-430.
- Blackburn, S. 2006. Antirealist Expressivism and Quasi-Realism.
- , 1993. *Essays in quasi-realism*. Oxford University Press.
- , 1993. *Morals and modals*.
- Bowler, D. M. 1992. “Theory of Mind” in Asperger's Syndrome. *Journal of Child Psychology and Psychiatry*, 33(5), 877-893.
- Byrne, A. 2011. *Knowing What I Want*.
- , 2011. *Knowing That I am Thinking*
- Chalmers, D. J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Churchland, P. M. 2013. *Matter and Consciousness*. MIT Press.
- , 1986. *Scientific Realism and the Plasticity of Mind*. Cambridge University Press.
- , 1981. Eliminative Materialism and the Propositional Attitudes. *The Journal of Philosophy*, 78(2), 67-90.
- Churchland, P. S. 2002. *Brain-wise: Studies in Neurophilosophy*. MIT Press.
- Corballis, M. C. & Sergent, J. 1988. Imagery in a Commissurotomized Patient. *Neuropsychologia* 26, 13-26.
- Crane, T. 2001. The Origins of Qualia: History of the Mind-Body Problem. T. Crane and S. Patterson. Eds., Routledge: 169-94.
- Davidson, D. 1963. Actions, Reasons, and Causes. *The Journal of Philosophy*, 60(23), 685-700.

Dennett, D. C. 1992. The Self as a Center of Narrative Gravity. In F. Kessel, P. Cole, and D. Johnson (eds) *Self and Consciousness: Multiple Perspectives*. Psychology Press.

Descartes, R. 2009. Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences. Reprinted, The Floating Press.

deVries, W. 2015. Wilfrid Sellars, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.) URL = <<http://plato.stanford.edu/archives/fall2015/entries/sellars/>>.

Eisenmajer, R., & Prior, M. 1991. Cognitive Linguistic Correlates of 'Theory of Mind' Ability in Autistic Children. *British Journal of Developmental Psychology*, 9(2), 351-364.

Finkelstein, D. 2003. Expression and the Inner. Harvard University Press.

Fodor, J. A. 1990. A Theory of Content and Other Essays. MIT Press, pp. 137-59.

—, 1987. Psychosemantics: The Problem of Meaning in the Philosophy of Mind. MIT Press.

—, 1975. The Language of Thought. Thomas Crowell.

Gazzaniga, M. S. 2014. The Split-Brain: Rooting Consciousness in Biology. *Proceedings of the National Academy of Sciences*, 111(51), 18093-18094.

—, 2000. Cerebral Specialization and Interhemispheric Communication. *Brain*, 123(7), 1293-1326.

Gallagher, S. 2004. Understanding Interpersonal Problems in Autism: Interaction Theory as an Alternative to Theory of Mind. *Philosophy, Psychiatry, & Psychology*, 11(3), 199-217.

Geach, P. T. 1965. Assertion. *The Philosophical Review*, 449-465.

Goldman, A. I. 1992. In Defense of the Simulation Theory. *Mind & Language*, 7(1-2), 104-119.

Gopnik, A. 1996. The Scientist as Child. *Philosophy of Science*, 485-514.

Gordon, R. M. 1992. The Simulation Theory: Objections and Misconceptions. *Mind & Language*, 7(1 - 2), 11-34.

—, 1986. Folk Psychology as Simulation. *Mind & Language*, 1(2), 158-171.

Hale, B. 1988. *Abstract Objects*. Oxford: Blackwell.

Hale, B. and Wright, C. 2009. The Metaontology of Abstraction In David Chalmers, Ryan Wasserman, and David Manley, eds, *Metaphysics: New Essays on the Foundations of Ontology*. Oxford University Press 178-212.

- , 2001. *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Clarendon.
- Happé, F. G. 1995. The Role of Age and Verbal Ability in the Theory of Mind Task Performance of Subjects with Autism. *Child Development*, 66(3), 843-855.
- Harman, G. 1990. The Intrinsic Quality of Experience. *Philosophical Perspectives*, 4, 31-52.
- Harris, P. L. 1992. From Simulation to Folk Psychology: The Case for Development. *Mind & Language*, 7(1-2), 120-144.
- Heal, J. 1986. Replication and Functionalism. In Jeremy Butterfield (ed.), *Language, Mind and Logic*. Cambridge University Press.
- Holton, R. 2004. Review of Wegner 2002. *Mind*, 113: 218–21.
- Hutto, D. D. 2007. The Narrative Practice Hypothesis: Origins and Applications of Folk Psychology. *Royal Institute of Philosophy Supplement*, 60, 43-68.
- Keeley, B. 2009. *The Early History of the Quale and its Relation to the Senses*. Routledge Companion to the Philosophy of Psychology, Routledge Press.
- Khlentzos, D. 2011. Challenges to Metaphysical Realism, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.) <<http://plato.stanford.edu/archives/sum2016/entries/realism-sem-challenge/>>.
- Kim, J. 2013. The Myth of Nonreductive Materialism. *The American Philosophical Association Centennial Series*, 641-658.
- Lewis, D. 1988. What Experience Teaches. Reprinted in *Papers in Metaphysics and Epistemology* (2009). Cambridge University Press, pp. 262–90.
- Libet, B. 2004. *Mind Time: The Temporal Factor in Consciousness*. Harvard University Press.
- , 2001. Consciousness, Free Action and the Brain, *Journal of Consciousness Studies*, 8: 59–65.
- , 1985. Voluntary Action. *Behavioral and Brain Sciences*, 8: 529–39.
- Matthews, R. J. 2007. *The Measure of Mind: Propositional Attitudes and Their Attribution*. Oxford University Press.
- Merricks, T. 2001. *Objects and Persons*. Oxford: Clarendon.
- Metzinger, T. 2004. *Being No One: The Self-Model Theory of Subjectivity*. MIT Press.

- Myers, R. E. 1956. Function of the Corpus Callosum in Interocular Transfer. *Brain* 79, 358-363.
- Myers, R. E. & Sperry, R. W. 1958. Interhemispheric Communication Through the Corpus Callosum: Mnemonic Carry-Over Between the Hemispheres. *Arch. Neurol. Psychiatry* 80, 298-303.
- Nagel, T. 1974. What is it Like to Be a Bat? *Philosophical Review*, 83: 435–450
- Nebes, R. 1973. Perception of Spatial Relationships by the Right and Left Hemispheres of a Commissurotomized Man. *Neuropsychologia* 11, 285-289.
- Ogden, C. K. and Richards, I. A., 1923. *The Meaning of Meaning*. Harcourt Brace & Jovanovich.
- Price, H. 2011. Expressivism for Two Voices. *Pragmatism, Science and Naturalism*. Academic Press. pp. 173–220
- , 2011. *Naturalism Without Mirrors*. Oxford University Press.
- , 2008. Two Readings of Representationalism. Descartes Lecture given at Tilburg August, 5
- Ross, L. 1977. The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process. In Berkowitz, L. *Advances in Experimental Social Psychology*
- Ruffman, T., Garnham, W., & Rideout, P. 2001. Social Understanding in Autism: Eye Gaze as a Measure of Core Insights. *Journal of Child Psychology and Psychiatry*, 42(8), 1083-1094.
- Ryle, G. 2009. *The Concept of Mind*. Routledge.
- , 1971. *Collected Papers*, volume 2. Hutchinson.
- Schiffer, S. 2003. *Things We Mean*. Oxford University Press.
- Schroeder, M. 2008. What is the Frege-Geach Problem?. *Philosophy Compass*, 3(4), 703-720.
- Shoemaker, S. 2000. Realization and Mental Causation. In *The Proceedings of the Twentieth World Congress of Philosophy* (Vol. 9, pp. 23-33).
- Searle, J. R. 1969. *Speech Acts: An Essay in the Philosophy of Language* (Vol. 626). Cambridge University Press.
- , 1962. Meaning and Speech Acts. *Philosophical Review*, 71 423-32.

Sellars, W. 1956. Empiricism and the Philosophy of Mind. *Minnesota Studies in the Philosophy of Science*, 1(19), 253-329.

Shoemaker, S. 1994. Self-Knowledge and "Inner Sense": Lecture I: The Object Perception Model. *Philosophy and Phenomenological Research*, 54(2), 249-269.
—, 1982. The Inverted Spectrum. *Journal of Philosophy*, 79: 357–81; page reference is to the reprint in Shoemaker 1984.

Sperry, R.W. 1961. Cerebral Organization and Behavior. *The American Association for the Advancement of Science, Science* Vol. 133, No. 3466, 1749 – 1757.

Squire, L. R. (Ed.). 2011. The History of Neuroscience in Autobiography (Vol. 7). Oxford University Press.

Stevenson, C. 1944. Ethics and Language. Yale University Press.
—, 1937. The Emotive Meaning of Ethical Terms. *Mind*, 46: 14–31.

Stitch, S. 1983. From Folk Psychology to Cognitive Science: The Case Against Belief. MIT Press.

Swanson, L. W. 2000. Cerebral Hemisphere Regulation of Motivated Behavior. *Brain Research*, 886(1), 113-164.

Tager-Flusberg H 2007. Evaluating the Theory-of-Mind Hypothesis of Autism. *Current Directions in Psychological Science* 16 (6): 311–315.

Tanney, J. 2013. Rules, Reason, and Self-Knowledge. Harvard University Press.
—, 2009. Rethinking Ryle: A Critical Discussion of The Concept of Mind
—, 2005. Reason-Explanation and the Contents of the Mind. *Ratio*, 18(3), 338-351.
—, 1995. Why Reasons May Not be Causes. *Mind & Language*, 10(1-2), 105-128.

Thomasson, A. L. 2015. Ontology Made Easy. Oxford University Press.
—, 2010. Ordinary Objects. Oxford University Press.
—, 2008. Phenomenal Consciousness and the Phenomenal World. *The Monist*, 91(2), 191-214.

Van Inwagen, P. 1990 Material Beings. Cornell University Press.

Wegner, D. M. 2004. Précis of The Illusion of Conscious Will, *Behavioral and Brain Sciences*, 27: 649–59.
—, 2002. The Illusion of Conscious Will. MIT Press.

Williams, M. 2010. Pragmatism, Minimalism, Expressivism. *International Journal of Philosophical Studies*, 18(3), 317-330.

Wittgenstein, L. 1991. *The Blue and Brown Books: Preliminary Studies for the 'Philosophical Investigation'*. Basic Blackwell.
—, 1953/58. *Philosophical Investigations* Oxford: Blackwell.