HOW BELIEFS ARE LIKE COLORS

Devin Sanchez Curry

A DISSERTATION

in

Philosophy

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2018

Supervisor of Dissertation

_____

Gary Hatfield
Adam Seybert Professor in Moral and Intellectual Philosophy
University of Pennsylvania

Graduate Group Chairperson

_____

Samuel Freeman, Avalon Professor of the Humanities, University of Pennsylvania

Dissertation Committee

Elisabeth Camp, Associate Professor of Philosophy, Rutgers University
Zoltan Domotor, Professor of Philosophy, University of Pennsylvania
Michael Weisberg, Professor and Chair of Philosophy, University of Pennsylvania

HOW BELIEFS ARE LIKE COLORS

COPYRIGHT

2018

Devin Sanchez Curry

This work is licensed under the
Creative Commons Attribution-
NonCommercial-ShareAlike 3.0
License

To view a copy of this license, visit
https://creativecommons.org/licenses/by-nc-sa/3.0/us/

Chapter 3 of this dissertation is derived, in part,
from "Beliefs as inner causes: the (lack of) evidence",
published in *Philosophical Psychology* on March 28th 2018,
available at https://doi.org/10.1080/09515089.2018.1452197

For Denise Roca Curry and David Cook Kemper Curry

Acknowledgements

Philosophers struggle to live good lives in worlds they find intelligible. For my mom—who thinks very highly of my writing while professing not to understand a word of it—these Socratic tasks are a breeze. She has no need for hoary metaphysics. My dad, in contrast, suggested I read Descartes's *Meditations* for a class project in 9th grade. I thought he was introducing me, his bookish kid, to the bookish thing he did for a living. I now know (fifteen years of formal schooling later) that my siblings and I had been talking philosophy and reading philosophical literature—Simak's *City*, Watterson's *Calvin and Hobbes* — with our dad throughout our childhood. This dissertation is dedicated to Denise and David Curry, for showing me how to live right. Thanks also to Galen and Tess, for being my constant companions in ethical education.

Nearly every good, bad, and weird idea I have had over the last seven years—including those that made it into this dissertation—has been tested in conversation with Amelia Brockman Curry. I am unspeakably lucky to look forward to a future centering on more conversations with Amelia. Thanks to her, and our cats Hobbes, Anouk, and Matisse, for always being there.

Gary Hatfield has been an ideal supervisor. He models passionate and careful scholarship, pursuing the questions that fascinate him across disciplinary borders. I cannot thank him enough for investing the same passion and care in me. Professor Hatfield takes his vocational roles admirably seriously. If I have been a good apprentice, it is because he took the time and thought to be a superb mentor.

I owe similarly deep thanks to my committee. I leave every meeting with Liz Camp charged up about talking philosophy, newly equipped with (exquisitely helpful) doubts about my habits of thinking. Despite his ability to classify every tree, Michael Weisberg never lets me miss the forest. I am regularly inspired by how Michael makes good on his boundless curiosity. Zoltan Domotor has an uncanny knack for seeing my arguments more lucidly than they had been represented in my head or on my page.

In addition to my official committee members, Lisa Miracchi and Dan Singer provided insightful and encouraging comments on large chunks of my dissertation. Susan Sauvé Meyer taught me to teach undergraduates. Karen Detlefsen has mentored me in ways that far exceed her obligations, challenging me (by advice and example) to go into the world and do the kinds of writing and teaching that make me happy. Before I came to Penn, Paul Davies and Kristin Andrews introduced me to the philosophical and psychological study of other minds.

More than any other factor, the comradery of my fellow graduate students made the Penn philosophy department a joyful place to work. I had the privilege of entering as a member of a dream cohort; Marie Barnett, Steve Esser, Daniel Fryer, Kurt Gerry, Nabeel Hamid, Karen Kovaka, Charles Phillips, and Jordan Taylor all became excellent philosophers and better friends. Marie, Nabeel, and Jordan, along with Ben Baker, Justin Bernstein, Louise Daoust, Carlos Santana, and Rob Willison, deserve special thanks for countless discussions about the nature of belief (and related topics). If space permitted, I would thank thirty more beloved grads by name. I already miss living alongside you all.

ABSTRACT

HOW BELIEFS ARE LIKE COLORS

Devin Sanchez Curry

Gary Hatfield

Teresa believes in God. Maggie's wife believes that the Earth is flat, and also that Maggie should be home from work by now. Anouk—a cat—believes it is dinner time. This dissertation is about what believing is: it concerns what, exactly, ordinary people are attributing to Teresa, Maggie's wife, and Anouk when affirming that they are believers. Part I distinguishes the attitudes of belief that people attribute to each other (and other animals) in ordinary life from the cognitive states of belief theoretically posited by (some) cognitive scientists. Part II defends the view that to have an attitude of belief is to live—to be disposed to act, react, think, and feel—in a pattern that an actual belief attributor identifies with taking the world to be some way. Drawing on scientific, scholarly, and literary sources of evidence, How Beliefs are like Colors *provides a framework for research on belief across the humanities and sciences of the mind.*

How Beliefs are like Colors

Table of Contents

Acknowledgements iv
Abstract vi
Table of Contents vii
List of Figures xi
Preface . . . . . xii–xx

Part I: Belief as Attitude, Belief as Cog 1–222

Chapter 1: A Rylean ancestry 2–59
1. Introduction 2
2. Dispositionalisms 4
2.1. Ryle 4
2.2. Baker 11
2.3. Schwitzgebel 17
3. Interpretivisms 20
3.1. Quine 20
3.2. Dennett 25
3.3. Davidson 36
3.4. Mölder 45
4. Ryleanisms 49

Chapter 2: Varieties of (theory of) belief 60–109
1. Introduction 60
2. Thirteen distinctions between varieties of belief 67
2.1. First distinction: implicit belief vs. explicit belief 68
2.2. Second distinction: binary belief vs. degrees of belief 70
2.3. Third distinction: system 1 belief vs. system 2 belief 72
2.4. Fourth distinction: dispositional belief vs. dispositions to believe 73
2.5. Fifth distinction: dispositional belief vs. occurrent belief 75
2.6. Sixth distinction: belief vs. assent 77
2.7. Seventh distinction: computational belief vs. algorithmic belief 79
2.8. Eighth distinction: personal belief vs. subpersonal belief 82
2.9. Ninth distinction: normative belief vs. natural belief 84
2.10. Tenth distinction: first-order belief vs. higher-order belief 87
2.11. Eleventh distinction: folk psychological belief vs. psychological belief 89
2.12. Twelfth distinction: manifest image belief vs. scientific image belief 95
2.13. Thirteenth distinction: attitude of belief vs. cognitive state of belief. 99
3. Ryleans and representationalists as bedfellows 100

4. Fifteen theoretical alternatives concerning the metaphysics of belief 103
5. The argument to come 108

Chapter 3: Beliefs as inner causes? 110–162
1. Introduction 110
2. Unpacking Davidson's dogma 113
2.1. ...beliefs as inner causes that produce behavior 113
2.2. The folk construe... 115
2.3. Davidson vs. Davidson's dogma 117
3. The folk craft of belief attribution 120
3.1. Cognitive scientific theories of belief attribution 121
3.2. A Rylean model theory of belief attribution 123
4. The (lack of) evidence from mindreading research 129
4.1. The empirical literature on lay belief attribution 129
4.2. A Rylean interpretation of the evidence about lay belief attribution 134
4.3. Mindreading researchers assume Davidson's dogma 136
5. The (lack of) evidence from attribution theory 140
5.1. The empirical literature on lay behavior explanation 140
5.2. A Rylean interpretation of the evidence about lay behavior explanation 146
5.3. Attribution theorists assume Davidson's dogma 147
6. Davidson vs. Ryle 149
6.1. Davidson's argument for his dogma 150
6.2. Ryle's anticipation of Davidson's argument 153
7. Conclusion 160

Chapter 4: How beliefs are like colors 163–222
1. Recap and introduction 163
2. Color 165
2.1. Against conflating perceivable colors with SSRs 165
2.2. Color relationalism 171
3. Belief 177
3.1. Against conflating attitudes with brain states 177
3.2. Against conflating attitudes with subpersonal computational states 180
3.3. Against conflating attitudes with personal purely functional states 185
3.4. How attitudes of belief are like perceivable colors 196
4. Evolution 199
4.1. Paramechanical teleofunctionalism about belief 200
4.2. Against conflating attitudes with teleofunctional cognitive states 206
5. Conclusion to Part I 217
5.1. The argument against conflating attitudes with cognitive states 217
5.2. The master argument for minimal Ryleanism about belief 219

Part II: The Metaphysics of Belief in Practice 223–380

Chapter 5: Beliefs as attributor-relative patterns of living 224–259
1. Introduction 224
1.1. Two wrinkles in my reformed Ryleanism 225
2. What beliefs are (and who they are for) 226
2.1. To believe is to live 227
2.2. To believe is to be disposed to act, react, think, and feel 229
2.3. To believe is to live in a pattern 232
2.4. To believe is to live in an identified pattern 233
2.5. To believe is to live in a pattern identified by a belief attributor 235
2.6. Attributors identify beliefs with believers *taking the world to be* some way. . . . 236
2.7. Attributors identify beliefs with believers taking the world to be *some way* . . 237
3. Seven common objections 240
3.1. Isn't 'taking the world to be some way' just belief? 241
3.2. Isn't belief attribution itself a kind of belief? 242
3.3. Aren't patterns of dispositions reducible to their categorical bases? 244
3.4. But aren't belief attributors *really* referring to categorical bases? 247
3.5. What about knowledge? 252
3.6. What about desire? 253
3.7. That's relativism! 254
4. Conclusion 259

Chapter 6: Styles of belief 260–307
1. Introduction 260
1.1. Ryle on Austen 260
2. Ryleans on variation between believers 267
2.1. Dennett and Baker 267
2.2. Schwitzgebel 269
2.3. Juliet and in-between belief 271
3. Improving on Schwitzgebel's conceptual machinery. 274
3.1. Cases of variance between believers 276
3.2. Styles of belief 279
3.3. Styles of believer 281
3.4. A (Ryle-on-Austen-ian) revision of Schwitzgebel's account 288
3.5. General-purpose stereotypes and believer-specific models 295
4. Lack of character? 299
5. Conclusion 307

Chapter 7: Relativistic Ryleanism, intersubjectivity, and norms 308–354
1. Introduction 308

1.1. Counseling the doubtful 308
1.2. Paramechanical denials of intersubjective indeterminacy 310
1.3. Rylean denials of intersubjective indeterminacy 311
2. Interpretivisms and norms 312
2.1. Davidson on intelligibility as the norm of belief attribution 313
2.2. Dennett on predictability as the norm of belief attribution 320
2.3. Mölder on community standards as the norms of belief attribution 324
3. Is there a normatively forceful constitutive norm of belief attribution? 327
3.1. Model-theoretic Ryleanism and the notion of an appropriate match 327
3.2. Diversity in models of belief 330
3.3. Variance in belief between attributors 335
4. A dilemma for Ryleans 338
4.1. Horn one: rejecting constitutive norms of belief attribution 339
4.2. Horn two: embracing a pluralistic normative Ryleanism 344
4.3. Relativistic Ryleanism and intersubjective indeterminacy 346
5. The messy metaphysics of belief in practice 347
5.1 What makes attributor-relative beliefs *beliefs*? 348
5.2. Can relativized attitudes of belief be objects of scientific inquiry? 352
6. Conclusion 353

Chapter 8: Brute believers and monkey mindreaders 355–380
1. Introduction 355
2. Anthropomorphism 356
2.1. Morgan's Canon 356
2.2. Anthropomorphism and attitudes of belief 365
3. Brute believers 367
4. Monkey mindreaders 371
5. Anthropomorphism redux 378
6. Conclusion: minding our differences 380

Bibliography 381–423

List of Figures

Figure 1: Thirteen distinctions between varieties of belief 68
Figure 2: Fifteen theoretical alternatives concerning the metaphysics of belief 104–105
Figure 3: Metamerism 166
Figure 4: Metamers viewed in another light 166
Figure 5: Intrasubjective variation in color vision 167
Figure 6: The dress: intersubjective variation in color vision 168
Figure 7: The dress photographed in daylight 169
Figure 8: Interspecific variation in color vision 170
Figure 9: Four desiderata for an account of attitudes of belief 251
Figure 10: A general-purpose stereotype and a style of belief 290
Figure 11: A very broad general-purpose stereotype and a style of belief 291

Preface

Teresa believes in God. Karl Ove believes he lacks what it takes to become a capital 'W' Writer. Lucinda believes she lost it. Maggie's wife believes that the Earth is flat, and also that Maggie should be home from work by now. Anouk—a cat—believes it is dinner time. This dissertation is about what believing is: it concerns what, exactly, we are thinking about Teresa, Karl Ove, Lucinda, Maggie's wife, and Anouk when we think that they believe.

I distinguish the *attitudes* of belief that people attribute to each other (and other animals) in ordinary life from the *cognitive states* of belief posited by (some) cognitive scientists. Cognitive states of belief may or may not exist. Regardless, attitudes of belief are tendencies of whole persons to live in patterns that interpreters identify with taking the world to be some way. This attributor-relative account of believing is uniquely able to account for the real but diverse phenomena tracked by researchers, scholars, novelists, and lay folk when they claim that so-and-so believes such-and-such. Believers base their senses of self, and self-worth, on their beliefs. We all attribute innumerable little beliefs to each other, weaving together our own idiosyncratic understandings of who the people in our lives are. We also praise, blame, trust, and fear people based on what they believe. Shared beliefs forge close friendships; divergent beliefs end romances. This dissertation provides a new framework for thinking about these core features of our social and personal lives.

Prefacing a book titled *The Analysis of Mind*, Bertrand Russell wrote that

"Psychology, theory of knowledge and metaphysics revolve about belief, and on the view we take of belief our philosophical outlook largely depends" (1921: xii). Russell did not declare belief to be the central aspect of mindedness, nor of the relationship between mind and world. Like Russell, I would be more inclined to treat *experience* as the central aspect of both mindedness and the mind-world relation. But I would rather not play that game at all; I am suspicious of any discussion of 'the mark of the mental', the supposed ur-aspect of mindedness that cleaves the minded from the mindless. Rationality, freedom of the will, the first-person point of view, intentionality, the appearance-reality distinction, the capacity to produce meaning, goal-oriented behavior, and consciousness have all been proposed as candidates for the mark. My suspicion is that this whole project rests on a mistake. Human and animal minds are hodgepodges of states, processes, mechanisms, capacities, stances, and phenomenal qualities. No single mark distinguishes the minded from the mindless.

Why, then, would our psychological and philosophical outlook revolve around belief, as Russell suggests? Here is one reason. Whereas experience may be a more central aspect of mindedness, belief is an extraordinarily salient dimension along which minds resemble and differ from one another. Nearly every human being across cultures develops the capacity at an early age to understand others in terms of what they believe. People see differences between minds—our own and others'—largely through the lens of differences between beliefs: differences in how people represent the world. These representational differences—differences in how people take the world to be—manifest

themselves in signature ways of acting, reacting, thinking, and feeling. The superstitious live as if black cats bring bad luck; Michael Jordan lives as if winning matters most; an unharried ice skater lives as if the river is frozen. People capture these distinctive ways of living by attributing beliefs. Experience may be metaphysically fundamental to mindedness, but the attribution of belief is epistemically fundamental to ordinary people's grasp on each other as uniquely sentient and sapient organisms.

There has, accordingly, been an enormous amount of social scientific research on what people believe. Nevertheless, belief has not emerged as a fecund object of inquiry in empirical cognitive science. In her excellent new book *Cognitive Gadgets*, the cognitive psychologist Cecilia Heyes dismisses beliefs as the purview of folk psychology: "the blend of wisdom and old wives' tales that we use to talk about the mind in everyday life" (2018: 10). Heyes celebrates the fact that cognitive scientists have instead focused on researching "cognitive mechanisms includ[ing] causal understanding, episodic memory, imitation, mindreading, normative thinking, and many more" (1–2). In contrast to beliefs and other mental phenomena individuated by folk psychology, these cognitive phenomena are individuated "by cognitive science, and, within the framework of cognitive science, different versions of a mechanism – different variants – can be distinguished according to what they do and how they do it; the kind of information they process, and the computations they use to process that information" (38). These distinctions make cognitive phenomena empirically tractable.

Like Heyes, I am a proponent of empirical cognitive science leading the way

when it comes to articulating theories of cognitive architecture. Moreover, I agree with Heyes that the proper objects of cognitive scientific investigation are capacities and kinds that are posited on the basis of experimentally collected behavioral evidence and which, in turn, prove empirically fecund. Now, the absence of 'belief' from Heyes's list of cognitive mechanisms should not be taken to imply that cognitive scientists do not posit belief as psychological kind. Most cognitive scientists do posit beliefs as functional cogs in cognitive systems. But their reasons for countenancing cognitive states of belief are mainly a priori and theoretical. Cognitive states of belief are not among those theoretical posits that have borne much empirical fruit in terms of unveiling the mechanisms of mind.

"Mindreading"—cognitive scientific jargon for the capacity to attribute belief—*is* on Heyes's list. Unlike believing itself, the capacities to construct, refine, and apply cognitive models of what people believe are all empirically robust cognitive phenomena, different versions of which can be distinguished according to the kind of information they process and the computations they use to process that information. Why would believing prove so elusive when the capacity to attribute belief has proven so experimentally fecund?

Perhaps because the relevant variety of believing is not actually a cognitive phenomenon. When people exercise their cognitive capacity to attribute belief, they are not ordinarily speculating about cognitive psychology. In my view, the wisdom and old wives' tales that feed human folk psychology do not aspire to posit another item for

inclusion in Heyes's list of cognitive mechanisms. Instead, when people attribute beliefs to somebody in ordinary life, they are trying to make sense of the believer's perspective, in light of what kind of person they are, without any concern for the deep functional structure of cognition.

Donald Trump believes that Mexico should pay for a U.S. border wall. Some configuration of his cognitive system is responsible for this belief. But when we attribute the belief in question to Trump, we are not trying to articulate that cognitive configuration. As Josh Marshall (2018) has recently argued, whatever is going on behind the scenes in Trump's cognitive system—whether he is mentally ill, for instance—"doesn't matter. If the entire psychiatric profession got together and examined Trump and pronounced him entirely free of any mental illness, his behavior wouldn't be any less whacked or dangerous in a President." What matters is that Trump manifestly lives—acts, reacts, thinks, and feels—in particular, particularly whacked, ways, which misrepresent important facts about the world. Attributing the belief that Mexico should pay for the border wall to Trump is a matter of grasping a pattern in how his actions, reactions, thoughts, and feelings (mis)represent the world, regardless of whether that pattern is the result of a disordered cognitive system. Quotidian practices of belief attribution do not require conflating such beliefs with objects of cognitive scientific investigation.

This idea—that the beliefs that people attribute to each other in everyday life are patterns of living rather than cognitive mechanisms—does not entail that cognitive

scientists (much less philosophers of mind) should lose interest in attitudes of belief. On the contrary, cognitive scientists have a large stake in explaining the proper objects of the cognitive capacity to attribute beliefs. The account of believing I develop in these pages draws amply on the extensive empirical literatures on belief attribution in social, cultural, and cognitive psychology, as well as work on belief and belief attribution in sociology, anthropology, and cognitive neuroscience. I also mix in a heavy dose of philosophical reflection and a sprinkling of insights gleaned from characterizations of believers in literary fiction.

Part I of the dissertation distinguishes the attitudes of belief that people attribute to each other (and other animals) in ordinary life from the cognitive states of belief theoretically posited by (some) cognitive scientists. Originally due to Gilbert Ryle, this distinction defuses much of the acrimony between philosophers concerned with understanding the objects of ordinary social cognition and philosophers concerned with understanding how cognitive systems function. Chapter 1 surveys the history of Rylean accounts of belief, Chapter 2 explores many varieties of belief and theories thereof, Chapter 3 marshals empirical evidence against the dogma that people ordinarily construe beliefs as inner causes, and Chapter 4 exploits an analogy between belief and color to argue against conflating attitudes with cognitive states.

Part II then develops and defends my relativistic Rylean view that to have an attitude of belief is to live—to be disposed to act, react, think, and feel—in a pattern that an actual belief attributor identifies with taking the world to be some way. Chapter 5

unpacks this definition and answers common objections, Chapter 6 investigates idiosyncratic styles of believing and develops the conceptual machinery required to adequately describe them, Chapter 7 argues that beliefs exist relative to the actual models wielded by individual belief attributors (rather than intersubjectively binding norms of interpretation), and Chapter 8 applies Ryleanism about belief to animal cognition research.

Ryle argued that "there is no place for 'isms' in philosophy" (1937: 153). The attendant irony of writing a dissertation defending (what I call) 'Ryleanism' is not lost on me. Ryle did not intend to establish a philosophical school. Nevertheless, Ryle's writings belie his proclamation; he plainly recognized that philosophical progress is often achieved by sussing out how strict adherence to isms—such as Ryle's favorite targets, Cartesianism and physicalism—falls short of wisdom. To that end, this dissertation can be read as the diagnosis of an ism. Part I focuses on what Ryleanism—the philosophy of belief adhered to by Gilbert Ryle, Daniel Dennett, Donald Davidson, Lynne Rudder Baker, Eric Schwitzgebel, and Bruno Mölder—gets right (and, incidentally, what "paramechanism" (Ryle 1949)—the view that attitudes of belief can be unproblematically conflated with cognitive states of belief—gets wrong). Part II then breaks free of the constraints set by traditional Ryleanisms, and develops a new relativistic alternative, inspired by but moving beyond Ryle's work.

In addition to the irony inherent in advocating (even a reformed) Ryleanism, I should fess up to a few shortcomings of the dissertation at the outset.

First, I provide an account of attitudes of belief, but no account of cognitive states of belief. I do argue that an account of cognitive states ought not be considered a prerequisite for an account of attitudes. Nevertheless, there remain open fascinating questions about the nature and reality of cognitive states, as well as their relationship to attitudes, that demand attention in future work. (Relatedly, I provide no account of original intentionality or content, as attitudes of belief feature only derived intentionality.)

Second, I deal with attitudes of belief almost entirely to the exclusion of other attributor-relative mental attitudes. As Donald Davidson wrote, "the word 'attitude' does yeoman service here, for it must cover not only permanent character traits that show themselves in a lifetime of behavior, like love of children or a taste for loud company, but also the most passing fancy that prompts a unique action, like a sudden desire to touch a woman's elbow" (1966: 686). A relativistic Ryleanism holds promise as an account of many attitudes beyond belief. Because each of these varied attitudes deserves special treatment, however, I focus on developing Ryleanism as an account of attitudes of belief alone.

Third, a complete account of the nature of attitudes of belief would elucidate the epistemic dimensions of belief. I say as little about them as possible; my relativistic Ryleanism is purposefully compatible with a wide range of positions in epistemology.

Fourth, I have not done the painstaking but invaluable ethnographic work required to lay bare the contours of actual belief attribution practices—that is, to reveal

how, precisely, particular attributors model particular beliefs when pursuing particular goals in particular real-world contexts. I provide a metaphysical account of attitudes of belief that situates such social scientific research within a sound theoretical framework. But I readily admit that it is a limitation of this account that it is built squarely on my armchair thought experiments and reading of the experimental literatures rather than more textured ethnographic evidence.

Metaphysicians of mind and cognitive scientists have too often erred in leaving the thick description (Ryle 1979; Geertz 1973) of mental qualities to ethicists, action theorists, cultural psychologists, and anthropologists. By my lights, theorists interested in what minds are and how they work must countenance aspects of mindedness that emerge within social contexts. The cognitive scientific task of determining how the machines inside our heads tick will be all the easier once we have a firmer grasp on which mental phenomena—like attitudes of belief—are attributor-relative human kinds (Hacking 1995), rather than purely functional cogs in cognitive systems (even when those cognitive systems are construed as embodied, extended, embedded, enactive, affective, or dynamically emergent). Moreover, one of the most edifying tasks for metaphysicians of mind lies in explaining how disparate mental phenomena—including cognitive functioning, reasoning, phenomenology, attributor-relative aspects of mind, and complex phenomena that straddle these borders—coordinate to form whole human and animal agents. I offer my relativistic Rylean account of attitudes of belief as an initial contribution to that larger, grander enterprise.

PART I: BELIEF AS ATTITUDE, BELIEF AS COG

If the seeming feuds between science and theology or between fundamental physics and common knowledge are to be dissolved at all, their dissolution can come not from making the polite compromise that both parties are really artists of a sort working from different points of view and with different sketching materials, but only from drawing uncompromising contrasts between their businesses. To satisfy the tobacconist and the tennis coach that there need be no professional antagonisms between them, it is not necessary or expedient to pretend that they are really fellow-workers in some joint but unobvious missionary enterprise. It is better policy to remind them how different and independent their trades actually are. Indeed, [the] smothering effect of using notions like depicting, describing, explaining, and others to cover highly disparate things reinforces other tendencies to assimilate the dissimilar and unsuspiciously to impute just those parities of reasoning, the unreality of which engenders dilemmas.

But you will not and should not be satisfied with this mere promise of a lifebelt. Can it be actually produced and thrown to us in the precise stretch of surf where we are in difficulties?

– Gilbert Ryle, *Dilemmas* (1954: 81)

Chapter 1: A Rylean ancestry

1. Introduction

Following Wilfrid Sellars, most philosophers consider beliefs to be theoretical entities.[1] According to the standard view, ordinary people wield a folk psychological theory that purports beliefs to be cogs in cognitive systems; this folk psychological theory will be either vindicated or refuted by empirical investigation into the mind. Most adherents to the standard view are either mind-brain identity theorists—who hold that (types or tokens of) beliefs are identical to (types or tokens of) brain states (Place 1956; Feigl 1958; Smart 1959; Armstrong 1968; Lewis 1966, 1972; Polger 2004; Shapiro 2018)—or role functionalists—who hold that beliefs are identical to functional roles contingently realized by brain states (Putnam 1967; Block & Fodor 1972; Millikan 1984; Fodor 1987; Dretske 1988; Jackson & Pettit 1990; Cummins 1996; Carruthers 2006; Burge 2010; Quilty-Dunn & Mandelbaum 2017).[2] Other adherents to the standard view are

---

[1] Sellars deserves credit for the standard view because his emphasis on the (pseudo-)theoretical nature of mental states was a catalyst for the rise of both functionalism and eliminativism in the philosophy of mind. However, Sellars inspired dispositionalism and interpretivism as well. Dennett (1987: 349), overstating the case, confesses that "almost no one cites Sellars, while reinventing his wheels with gratifying regularity." For example, Sellars stressed that "while it is, I suppose, noncontroversial that when the total scientific picture of man and his behavior is in, it will involve some identification of concepts in behavior theory with concepts pertaining to the functioning of anatomical structures, it should not be assumed that behavior theory is committed *ab initio* to a physiological identification of all its concepts, — that its concepts are, so to speak, physiological from the start" (1956: 185). As will become clear, the rejection of this assumption—as well as the Sellarsian assumption that folk psychological concepts are functional from the start—is central to theories of belief in the Rylean lineage.

[2] Functionalists disagree about whether to characterize the relevant functional roles as subpersonal computational roles, person-level functional dispositions, or etiological

eliminativists, who wager that the folk psychological theory that posits beliefs will be discredited as cognitive psychology and neuroscience advance, and that beliefs therefore do not exist (P.M. Churchland 1979, 1981, 1993; Stich 1983; P.S. Churchland 1986; Ramsey, Stich & Garon 1990; Chemero 2009; Rosenberg 2016).

Other philosophers deny that beliefs are theoretical entities: they neither consider beliefs to be cogs in cognitive systems nor conclude that beliefs therefore do not exist. In this chapter, I survey the positive views of these heterodox philosophers of belief. These philosophers include dispositionalists—who argue that to believe is to have an appropriate pattern of dispositions (without theoretical prejudice as to the functional role those dispositions or their psychological underpinnings might play in the believer's cognitive system)—and interpretivists—who argue that to believe is to be aptly interpretable as believing.

In the next two sections, I interpret the writings on belief by three prominent dispositionalists and three prominent interpretivists, respectively. In chronological order, the philosophers I review are Gilbert Ryle, Daniel Dennett, Donald Davidson, Lynne Rudder Baker, Eric Schwitzgebel, and Bruno Mölder. In the third section, I argue that all six of these philosophers have proposed (more or less idiosyncratic) versions of the same theory, which I call 'Ryleanism about belief'. Moreover, I argue that, in principle, if they are to be true alternatives to the standard view, dispositionalism needs interpretivism and *vice versa*. There are no plausible dispositionalisms or interpretivisms

---

teleofunctional roles. I will discuss these varieties of functionalism at length in Chapter 4.

that are not Ryleanisms. This analysis sets the stage for the remainder of Part I, in which I defend the generic form of Ryleanism to which all dispositionalists and interpretivists (at least implicitly) subscribe.

In this chapter, I review and synthesize 60-odd years of a (sometimes hermeneutically challenging) strand of work on the nature of belief. In order to accommodate this ambitious scope, the chapter is modest in other respects. Most notably, it contains no sustained arguments for Ryleanisms, much less against other views.[3] It serves only to provide the history and set out the general structure of Ryleanisms in order to clarify how an account of belief can simultaneously rival the mind-brain identity theory, various functionalisms, and eliminativism, insofar as these theories all subscribe to the standard view that the beliefs people attribute to each other in everyday life are theoretical entities.

2. Dispositionalisms

2.1. Ryle

With the 1949 publication of *The Concept of Mind*, Gilbert Ryle launched the dominant program in latter-half-of-the-20th-century philosophy of mind. Whitewashing important exceptions, the post-Ryle zeitgeist turned away from the first half of the century's concern with the epistemology of perception and towards concern with the

[3] I have also reserved my criticisms of the particular versions of Ryleanism put forward by Ryle, Dennett, Davidson, Baker, Schwitzgebel, and Mölder for Part II, though I flag my biggest disagreements in footnotes.

metaphysics of mental states. This program was not, by and large, positively Rylean. (The Rylean strand I follow in this chapter was but a tributary of the great river of metaphysics of mind that Ryle set in motion.) But, with Ryle, much of the program self-identified as staunchly anti-Cartesian.[4]

According to the first sentence of *The Concept of Mind*, Descartes's substance dualism—which Ryle went on to refer to "with deliberate abusiveness" as "Descartes' myth" and "the dogma of the Ghost in the Machine" (1949: 15–16)—was "so prevalent among theorists and even among laymen that it deserves to be described as the official theory" (11). Ryle was not a fan of Descartes's myth, but he did think there was one good thing about it. Unlike physicalist philosophers like Hobbes, Descartes and his followers recognized that "the mental could not be just a variety of the mechanical" (19). However, Ryle lamented that Descartes's official theory rightly "realized that the problem [of accounting for the mind] was not one of mechanics and [wrongly] assumed that it must therefore be one of some counterpart to mechanics" (22). In this way "the Cartesian myth does indeed repair the defects of the Hobbist myth only by duplicating it" (330). In Ryle's estimation, substance dualism sensibly rejected the unworkable idea that minds were machines only to replace it with the ludicrous "paramechanical" idea

[4] The post-Rylean program is only unproblematically termed 'anti-Cartesian' if we accept most of its practitioners' simplistic understanding of Descartes's dualism. Hatfield (2017) provides a more balanced and precise interpretation of Descartes's views on the relationship between bodies and minds; I have produced my own survey of the interpretive options on the Cartesian mind-body union in the context of a discussion of Descartes's views on the neurophysiology and psychology of memory (Curry 2018: Section 4.3).

that they were ghostly machines that operated bodily machines from the inside.

Ryle's book is often credited with securing the demise of substance dualism. Ironically, Ryle's critique cleared the path for a physicalist variety of functionalism—the only metaphysics of mind that Ryle deemed less tenable than substance dualism—to emerge as the official theory in its stead.

Ryle meant his own theory of mind to avoid both the shallowness of physicalism and the excesses of substance dualism. Ryle sought this middle ground by arguing that both sorts of theory made a category mistake—like that involved in visiting the library, museum, admissions office, and philosophy department, and then asking to visit the university (16). According to Ryle, both physicalists and dualists mistakenly categorize minds as entities, like mountains, people, and water bottles. Ryle proposed that mental terms instead refer to a variety of patterns manifest in how people live. Some mental terms refer to the manners in which people act, react, think, and feel (135–153). Others refer to the success (or failure) of certain behaviors (1954: 93–110). Yet others refer to the capability to behave (or behave well) (1949: 25–61). Finally, some mental terms, like 'believes that', mean that a person "is prone to do and feel certain things in situations of certain sorts" (116). Ryle thus pioneered contemporary dispositionalism about belief.

For Ryle, to say that somebody believes is not to posit a cognitive state of belief—a sentence in the Language of Thought, say—that contributes to the production of their behavior. A belief is neither a cog nor an episode; it is a "determinable disposition" (118), "the exercises of which are indefinitely heterogeneous" (44). Determinable

dispositions are to be distinguished from determinate – or single-track – dispositions, "the actualisations of which are nearly uniform" (43).[5] Ryle calls "the brittleness of glass" a single-track disposition: "to be brittle is just to be bound or likely to fly into fragments in such and such conditions" (43). But it would be wrong to try to make belief fit this model and

> postulate that, for example, a man who believes that the earth is round must from time to time be going through some unique proceeding of cognising, 'judging', or internally re-asserting, with a feeling of confidence, 'The earth is round'. In fact, of course, people do not harp on statements in this way. (1949: 44)

No, a belief is a determinable disposition; or, to put the same point differently, a belief is a pattern of dispositions. For even if somebody did go around declaring the earth to be round, "we still should not be satisfied that they believed that the earth was round, unless we also found them inferring, imagining, saying and doing a great number of other things as well" (44). In particular, "believe ... is a tendency verb" (133): to say that someone believes is to say that they tend to engage in a certain range of behaviors, think a certain range of thoughts, and feel a certain range of feelings.[6] 'Belief' is also a "motive

---

[5] Ryle used the term 'determinable dispositions' but (to my knowledge) never the term 'multi-track dispositions', despite the attribution of the latter term to him by Dennett's Introduction to the *Concept of Mind* (Ryle 1949: vii) and other sources (Fara & Choi 2012).

[6] The popular miscategorization of Ryle as a 'logical behaviorist' notwithstanding, he numbered "feeling" (1949: 92, 1954: 59), "inferring, imagining" (1949: 65), "thinking, pondering, reflecting,

word" (133): to say that someone believes something is to answer the question why they—what motivates them to—do those things, think those thoughts, and feel those feelings.

Following Bertrand Russell (1910), philosophers usually think of beliefs as 'propositional attitudes': relations thinkers bear to propositions. In particular, believing is typically supposed to be the relation of a taking a proposition to be true. For Ryle, this view "though not far wrong, is too narrow" (134). Consider a skater who believes he skates on thin ice: "certainly to believe that the ice is dangerously thin is to be unhesitant in telling oneself and others that it is thin, in acquiescing in other people's assertions to that effect, in objecting to statements to the contrary, in drawing consequences from the original proposition, and so forth" (135). But expressing an attitude towards a proposition is not nearly sufficient for having a belief. To believe the ice is thin "is also to be prone to skate warily, to shudder, to dwell in imagination on possible disasters and to warn other skaters" (135). Indeed, assent to a proposition is not even necessary for belief; "however often and stoutly a skater avers to us or to himself, that the ice will bear, he shows that he has qualms, if he keeps to the edge of the pond, calls his children away from the middle, keeps his eye on the life-belts or continually speculates what would happen, if the ice broke" (45).

---

and the like, namely … what Rodin's *Le Penseur* looks as if he is absorbed in" (1979: 65), among the passivities and activities towards which believers are disposed. Sellars's famous Rylean ancestors (1956: §48) are in this respect inaptly named. Ryleans worthy of the name do not need a genius to come around to show them how to notice their sensations and make their quintessentially public speech private; Ryle already allowed for Jones's innovation.

Ryle unpacked his notion of 'determinable dispositions'—including beliefs—via the metaphor of 'inference tickets'. He wrote that if claims about the existence of determinable dispositions "are true, they are satisfied by narrated incidents" (125). For example:

> 'John Doe has just been telephoning in French' satisfies what is asserted by 'John Doe knows French', and a person who has found out that John Doe knows French perfectly needs no further ticket to enable him to argue from his having read the telegram in French to his having made sense of it. Knowing that John Doe knows French is being in possession of that ticket, and expecting him to understand this telegram is traveling with it. (1949: 125)

This example concerns the attribution of knowledge, which denotes certain dispositional *capacities*, such as the ability to parse French sentences. The attribution of belief, which instead denotes dispositional *tendencies*, is an inference ticket as well. By justifiably interpreting John Doe as believing that his friend Pierre speaks only French, the attributor is in possession of a ticket that licenses (among other inferences) the inference that John Doe will greet Pierre with a 'bonjour' instead of a 'howdy-do'. By justifiably interpreting the skater as believing he is on thin ice, the attributor is in possession of the ticket that licenses inferences to the conclusions that he will behave warily and experience inward terror. Beliefs—qua patterns of dispositions—exist insofar as their

attribution provides belief attributors with inference tickets. Insofar as beliefs are defined by their functional roles, the function of beliefs is to license inferences made by belief attributors, rather than to power cognitive machines.

A final dimension of Ryle's dispositionalism that deserves attention is his insistence that attributions of belief, qua inference-tickets, should not be construed "as asserting extra matters of fact" (119). The dogma of the ghost in the machine was wrongheaded precisely because it held beliefs and other mental phenomena to be extra "things existing, or processes taking place, in a sort of limbo world" (120), over and above the things that exist and the processes that take place in the physical world. For Ryle, saying that the physical world contains people with beliefs (as well as the objects and forces posited by physics) is not to say that there is anything in the world extraneous to the objects and forces posited by physics. On the contrary, "the truths of fundamental physical theory are … truths about anything whatsoever in the world" (1954, 74). Talking about belief is merely an enormously useful (and equally true) way of describing the very same world that is described differently by physicists—and described differently still by artists and accountants and neuroscientists. Belief talk does not compete with ledger-book talk or neuron talk. To describe someone in terms of their beliefs—and thus in terms of their tendencies and motives—serves a compatible but incommensurable cross-purpose to describing them in terms of their neurons.[7] To say

[7] Ryle's (1954) *Dilemmas* provides an in-depth look at cases in which philosophers have taken different ways of describing the world to be incompatible though, according to Ryle, they are not.

that a person believes something, for Ryle, is to say nothing more than that we are licensed to infer that the believer is prone to do, think, and feel certain things in situations of certain sorts. We are licensed to these inferences because the believer is truly intrinsically disposed to live in the attributed pattern. Nevertheless, ordinary belief attribution entails *nothing* about the cognitive capacities underlying the pattern of dispositions in question: that is a topic for the psychologists and neuroscientists, not the lay belief attributor.

In the decades following the publication of *The Concept of Mind*, original dispositionalist theories of belief have been introduced and defended by Price (1969), Audi (1972, 1994), Barcan Marcus (1990, 1995), Searle (1992), Baker (1987, 1995, 2003), and Schwitzgebel (2001, 2002, 2010, 2013, 2017), among others. I turn now to reviewing Baker and then Schwitzgebel's especially systematic and influential dispositionalisms.

2.2. Baker

Lynne Rudder Baker is not uncontroversially classified as a dispositionalist. Indeed, she declares herself to be both anti-Ryle and anti-dispositionalism. Baker takes herself to be anti-Ryle because "against Ryleans … I hold that belief explanations are causal explanations" (1995: 28), and anti-dispositionalist insofar as "we think of dispositions in terms of stimulus/response pairs requiring a single intrinsic causal basis" (2003: 191). In other words, Baker rejects the ideas, which she associates with Rylean and dispositionalist thinking respectively, that beliefs are not causally explanatory, and that

beliefs are single-track dispositions which are causally derivative of physical states of the brain.

Neither of these purported antagonisms holds up under scrutiny. Baker rejects reductionist dispositionalisms (e.g. Stalnaker 1984) that rely on the popular Lewisian (1997) metaphysics of dispositions, but she is a steadfast dispositionalist of Ryle's ilk. As we have seen, Ryle argued that beliefs are determinable dispositions and rejected the idea that dispositionalists must posit intrinsic causal bases of belief. Baker readily admits that "if we understand dispositions in a bare, schematic way – the meaning of the attribution of a disposition is exhausted by certain counterfactuals – then beliefs in this broad sense are dispositions" (2003: 195). Her main goal, like Ryle's, is to show how we can give up substance dualism while maintaining that beliefs "are global states of whole persons," rather than subpersonal states that are either identical to or categorically based in "(perhaps distributed) brain states" (187).

What, then, of Baker's purported opposition to Ryle over whether belief explanations are causal explanations? In brief, the manner in which Baker takes belief to be causally explanatory is not the manner in which Ryle denied belief to be causally explanatory. For Baker, a causal explanation "affords control over phenomena of the type explained. When we can produce or prevent a phenomenon at will, we know that we have found a cause" (1995: 122). She proposes the following test which "purports to provide … a sufficient condition for a causal explanation" (121): an occurrence of one phenomenon causally explains another "if: (i) If [the purported cause] had not occurred

in [the context], then [the purported effect] would not have occurred" and "(ii) given that [the purported cause] did occur in [the context], an occurrence of [the purported effect] was inevitable" (122).[8] Having argued for this view of how beliefs cause behaviors, thoughts, and feelings, Baker cites Ryle's view that "motives [like belief] are not happenings and are not therefore of the right type to be causes" (Ryle 1949: 113), and claims that "such statements at least suggest that, whereas I regard belief explanations typically to be causal explanations, Ryle would not" (Baker 1995: 237).

Baker is correct that Ryle denied that beliefs are causes in an important sense: they are not events that causally produce other events. But Baker does not take beliefs to be causes in *that* sense either, and Ryle would have agreed that beliefs are causally explanatory according to Baker's test: we can explain the skater's having skirted (rather than skated over) the ice by referencing the facts that (i) if he hadn't believed that the ice was thin, then he wouldn't have skirted the ice, but (ii) given that he did believe that the ice was thin, it was inevitable (in context) that he would skirt the ice.[9] Indeed, even as Ryle denied that motive words like belief designate producing causes, he scooped Baker in assigning them the important explanatory function of answering *why* somebody behaves as they do (Ryle 1949: 88–89). With the answer *why* in hand, (insofar as we can control what people believe) belief attributors can (to that extent) produce or prevent the

---

[8] Baker's account of dependence causation bears strong similarities to Woodward's (2003) influential interventionism about causation.

[9] I delve deeper into the two varieties of causation, as well as the case for the Rylean view that beliefs are dependence rather than producing causes, in Chapter 3.

behaviors motivated by people's beliefs.

Baker's theory of belief, which she terms 'Practical Realism', is a largely (if unwittingly) Rylean brand of dispositionalism. Practical Realism is the theory that "whether a person *S* has a particular belief (individuated by a 'that' clause in its attribution) is determined by what *S* does, says, and thinks, and what S *would* do, say, and think in various circumstances, where 'what *S* would do' may itself be specified intentionally" (1995: 154–155). The counterfactuals here are especially important. As Baker stresses, "Practical Realism allows that one may have a belief even if one never manifests it in overt behavior" (155). Thus, "one has a belief that *p* by virtue of there being a conjunction of relevant true counterfactuals that mention a range of circumstances in which the believer would perform a range of intentional actions – thinkings, doings, and sayings" (187). Working with the broad understanding of dispositions that Baker (like Ryle) urges, to believe is to have a particular pattern of behavioral, verbal, and cognitive dispositions.

What fixes the relevant pattern of dispositions? To answer that question, Baker situates her theory of belief in "an equally comprehensive metaphysics" (220). Challenging Paul Churchland's claim that "our commonsense conceptual framework can be seen as as speculative and as artificial as any overtly theoretical system" (Churchland 1979: 2), Baker counters that "the commonsense conception is not theoretical in the same way that the sciences are; yet it is a reliable source of truth" (220). According to Baker, again echoing Ryle, "the commonsense conception functions as a

cognitive background for … the world of human affairs – the world in which some professors have tenure, Germany is reunified, recycling centers are opened, budgets are cut, and all the rest – this world is no less real, no less noble, than the austere world of basic particles whose duration is too short to be measured" (223, 228). Our everyday understanding of this world of human affairs refers inexorably to beliefs: the belief that Professor X is a prodigious publisher, the belief that the fall of the Soviet Union falsifies Marxism, the belief that our money should be allocated to going green rather than employing middle school art teachers. For Baker, common sense is the arbitrator of non-theoretical phenomena. Common sense conceives humans to possess myriad beliefs, and conceives the possession of beliefs to amount to having the right patterns of dispositions.

Baker does not take the commonsense framework to be relativistic. She writes that while "a commonsense conception is a conception of reality that one learns in learning a natural language" (221), it nevertheless "is not restricted to some particular outlook that may vary from culture to culture. Rather, it provides a common background against which differences among cultures become visible" (222). Baker thus insists that there is an objective, intersubjectively accessible fact about whether or not somebody has a particular belief. She grounds the objectivity of belief in a notion of recognition-independence, which she distinguishes from mind-independence. To be mind-independent is to "exist in some possible world in which there are no concepts of anything" (231). To be recognition-independent, on the other hand, is to "be instantiated

independently of anyone's awareness" (232). Beliefs are mind-dependent: there are no beliefs in worlds that lack minds with access to the commonsense conception. But this mind-dependence does not impugn the objectivity of belief; as Baker points out, "carburetors … cannot be instantiated in worlds without minds" (232) either. There are nevertheless objective facts about whether or not something is a carburetor. Baker claims that this objectivity obtains because "the property of being a carburetor is recognition-independent. Something is a carburetor even if no one is ever aware that it is" (232–233). Likewise, in order to believe that *p*, a believer does not have to be interpreted as believing that *p*. The believer only has to fit the commonsense dispositional profile for the belief, whether or not anybody ever recognizes that fact.[10]

To summarize, Baker's view is that believing is counterfactually doing, saying, and thinking in a certain pattern, as objectively specified (mind-dependently but recognition-independently) by the commonsense conception. Baker's account of the commonsense conception is sketchy: she gives no details about the particular dispositions that are supposed to be associated with particular beliefs. Ryle's original dispositionalism was similarly slim on specifics. Eric Schwitzgebel intends his recent version of dispositionalism to provide the means to fill in these details. As he puts it, "the focus of the present account is on what tended to be only an afterthought to the thrill of exorcising the dualist's ghost-in-the-machine [as well, we might add, as the

[10] I will criticize Baker's view that this kind of recognition-independence precludes relativism about belief in Chapters 5 and 7.

physicalist functionalist's cog-in-the-machine]: a description of the conditions under which particular beliefs can properly be attributed to human beings" (Schwitzgebel 2002: 258). Schwitzgebel proposes a version of dispositionalism that is sensitive to idiosyncratic cases, especially cases in which it is not entirely clear what (or even whether) somebody believes.

2.3. Schwitzgebel

Citing Annette Baier (1985) and John Locke (1690), Schwitzgebel refers to beliefs as "temporary or habitual posture[s] of the mind" (2013: 3). Schwitzgebel's theory is that "to believe that P … is nothing more than to match to an appropriate degree and in appropriate respects the dispositional stereotype for believing that P" (2002: 251). Schwitzgebel defines a stereotype as "a cluster of properties we are apt to associate with a thing, a class of things, or a property" (250), and "a *dispositional stereotype* [as] a stereotype whose elements are dispositional properties" (251). The dispositional properties that Schwitzgebel takes belief attributors to be apt to associate with beliefs include not only behavioral dispositions but also "*phenomenal* dispositions, dispositions to have certain sorts of conscious experiences", and *cognitive* dispositions, "dispositions to enter mental states that are not wholly characterizable phenomenally" (252). Schwitzgebel characterizes dispositional properties in much the same way as Baker: "Dispositions can be characterized by means of conditional statements of the form: If condition C holds, then object O will (or is likely to) enter (or remain in) state S" (252).

Putting these definitions together, Schwitzgebel holds that beliefs are identical to the cluster of behavioral, phenomenal, and cognitive dispositions that belief attributors are apt to associate therewith.

Schwitzgebel stresses that believing is not always an all or nothing affair.[11] The skater might believe, 100%, that he is on thin ice, and possess all of the dispositions attributors are apt to associate with that belief. Or the skater might lack the belief entirely, possessing none of the stereotypical dispositions as he skates obliviously forth and the ice cracks beneath his blades. But "between these two extremes is a wide range of cases in which the subject has some but not all the dispositions in the stereotype" (252). Many skaters are wary, though indisposed to proclaim that the ice is thin. These skaters, who fit the dispositional stereotype for believing the ice is thin more or less well, can be described as being in what Schwitzgebel describes as "an in-between state of belief" (2001: 76). Schwitzgebel has offered a wealth of cases of in-between believing (2001: 76–78; 2002: 260–263; 2013: 16–19). Consider Ellen, who has spent four years studying Spanish in high school. Ellen is simultaneously disposed "to assent to the claim that all Spanish nouns ending in 'a' are feminine" with confidence, and to use the nouns 'anarquista' and 'bolchevista' "as masculine when the situation demands" (2002: 260). Does Ellen believe that all Spanish nouns ending in 'a' are feminine? Schwitzgebel

[11] According to the Bayesian program, beliefs should be characterized in terms of degrees of confidence between 0 (no confidence) and 1 (complete confidence). While Schwitzgebel agrees with the Bayesians that beliefs are not simply an on/off affair, he denies that thinking in terms of degrees of belief suitably explains cases of in-between believing (Schwitzgebel 2001: 78–80).

suggests that the best answer is: sort of. Ellen is best characterized as being in-between believing and not believing, since she possesses some but not all of the dispositions stereotypically associated with the belief in question.[12]

Schwitzgebel quickly glosses over one of the innovations of his dispositionalism. Unlike Ryle and Baker's, as well as the interpretivisms of Dennett, Davidson, and Mölder, Schwitzgebel's dispositionalism allows for relativism about the pattern of dispositions associated with a given belief. He admits that "not every person will be apt to associate exactly the same dispositional properties with any given belief – no more than everyone would agree about the stereotypical properties of a chair", but quickly assures the reader that "we can agree to a substantial extent about what are and are not typical chairs. A similar general but imperfect concord is available in the case of belief." Schwitzgebel therefore explicates his dispositionalism "as though there were one stereotype for every belief, though strictly speaking this is not true" (251). He does further damage control in an endnote, writing that "the relativism about belief that may be engendered by my account is ontologically superficial – it is a consequence only of variations in people's shorthand ascriptions and does not reflect real differences in the subjects themselves" (271).[13] Indeed, Schwitzgebel takes his entire project to be ontologically superficial. While coy about his relativism, Schwitzgebel is unabashed

---

[12] I suggest an alternative way of understanding this style of case in Chapter 6.

[13] I criticize Schwitzgebel's wariness about relativism—and related recourse to claims of ontological superficiality—in detail in Chapters 5–7. My own version of Ryleanism is ontologically serious and unabashedly relativistic, though not in a manner that threatens the objective assessibility of belief attributions.

about the generally anti-ontological spirit of his dispositionalism. He notes that "the metaphysically inclined may wish to ask whether, useful or not, the present account accurately describes what belief *really* is" (270). Schwitzgebel shrugs in response. He claims only to be concerned that his account is helpful in our attempt to understand the phenomenon of belief, in all of its quotidian messiness.

The three dispositionalisms reviewed in this chapter are differently motivated elaborations of the same basic theory. For Ryle, whether a pattern of dispositions is appropriately identified with a belief is determined by the logic of thought and talk about belief. For Baker, the appropriateness is arbitrated by the folk wisdom inherent in the commonsense conception. For Schwitzgebel, it comes down to individual interpreters' stereotypes—though he often avers to community standards for ease of exposition (2013: 80). Regardless, for Ryle, Baker and Schwitzgebel, to believe is to possess an appropriate pattern of dispositions to act, react, think, and feel.[14]

3. Interpretivisms

3.1. Quine

Daniel Dennett and Donald Davidson's independently developed interpretivisms are best introduced by way of their shared source: Willard Van Orman's

[14] Schwitzgebel speaks of clusters of determinate dispositions, whereas Ryle speaks of single determinable dispositions. Baker (2003: 194–195) rightly notes that there is no substantive difference between these two ways of talking. From here on out, I use both locutions interchangeably (with each other and my own preferred term: 'patterns of dispositions'.

Quine's exposition of the indeterminacy of meaning.[15] Quine exposed the indeterminacy of meaning through a thought experiment in which a linguist attempts "*radical* translation, i.e., translation of the language of a hitherto untouched people [where] all help of interpreters is excluded" (1960: 28). Quine asked whether the radical translator could determine precisely what was meant by each and every utterance in the erstwhile untranslated language.

For example, when "a rabbit scurries by [and] the native says 'Gavagai'", could the radical translator figure out the exact meaning of 'Gavagai' through further investigation? Quine thought not. The radical translator might be able to rule out lots of possible meanings: 'Gavagai' probably does not mean 'boat', 'quark', 'beautiful day' or 'ennui'. But she could not determine if Gavagai meant precisely 'rabbit', as opposed to 'brief temporal segments of rabbits' or 'all and sundry undetached parts of rabbits'. Thus, Quine wrote that "when from the sameness of stimulus meanings of 'Gavagai' and 'Rabbit' the linguist leaps to the conclusion that a gavagai is a whole enduring rabbit, he is just taking for granted that the native is enough like us to have a brief term for rabbits and no brief general term for rabbit stages or parts" (1960: 52). The linguist cannot determinately translate 'Gavagai' because she cannot assume that her subjects understand and describe the world in terms of the same basic objects as she does. Quine variously called this phenomenon "ontological relativity" (1969) and the "inscrutability

---

[15] Dennett reflects that "I'm what you get when you cross Quine with Ryle and add some cognitive science" (Dennett 2008b; cf. Dennett 1994: 365; Pyle 1999: 69). Davidson, meanwhile, dedicated his *Inquiries into Truth and Interpretation* to Quine, "without whom not." (2001: v).

of reference" (1981).

Quine took ontological relativity to be the "trivial and indisputable" cousin of "the serious and controversial thesis of indeterminacy of translation" (1992: 50). The fact of ontological relativity makes absolutely no difference to the pragmatic uses to which the radical translator might put her translation. She knows that, whatever 'Gavagai' means, it refers to (what she thinks of as) the rabbit. The radical translator can easily predict and explain what her subject does based on what her subject says, even though she cannot know for sure whether 'Gavagai' means 'rabbit' or 'rabbit time slice' or any number of other (conditionally) co-referring expressions.

The important thesis of indeterminacy, for Quine, is not ontological relativity but "full or holophrastic indeterminacy of translation" (1992: 50). Quine held that in many situations the radical translator would fail not only to determine the meaning of individual terms, but also to determine what her subject meant by whole sentences. Imagine a situation in which the radical translator is at a loss to say whether the foreigner believes (A) that the rabbit is conscious or (B) that the philosophical zombie rabbit lacks consciousness but behaves like a conscious rabbit.[16] Unlike the ontologically

[16] This silly example is not a perfect illustration of what Quine had in mind. What Quine had in mind was that (a) and (b) were beliefs about physics that were equally well accounted for by the empirical evidence—Quine held that scientific theories are always underdetermined by the evidence, a thesis that he sometimes referred to as the "third indeterminacy of meaning" (1990). However, Quine himself admitted that "unlike indeterminacy of reference, which is so readily illustrated by mutually compensatory adjustments within the limits of a single sentence, the full or holophrastic indeterminacy of translation draws too broadly on a language to admit of factual illustration" (1992: 50).

relative comparison of 'rabbit' and 'rabbit time slice', (A) and (B) directly and substantively conflict. But there is still no way for the radical translator to distinguish between the two hypotheses. Quine wrote that "the question whether [in some such situations] the foreigner *really* believes A or believes rather B, is a question whose very significance I would put in doubt" (1970: 180–181). Similarly, "it will often happen also that there is just no saying whether to count an affirmation of propositional attitude as true or false, even given full knowledge of its circumstances and purposes" (1960: 218). According to Quine, it might be impossible for the radical translator to say whether her subject believes that the rabbit is conscious, because there is no determinate fact of the matter about what her subject really believes.[17]

Quine sometimes flirted with Ryleanism about "the essentially dramatic idiom of the propositional attitudes" (1960: 219). He wrote that "when a naturalistic philosopher [such as himself] addresses himself to the philosophy of mind, he is apt to talk about language" and especially "meanings … those very models of mental entities" (1969: 26). Allying himself with John Dewey's claim that "meaning … is not psychic existence; it is

[17] Dennett's example of full intersubjective indeterminacy—in which Brown and Jones disagree over whether Ella desires to kill herself—is discussed later in this chapter and in Chapter 4, and my own example is discussed at length in Chapter 7. While both Dennett and I countenance (at least the possibility of) full indeterminacy, we both disagree with Quine that "it will often happen also that there is just no saying whether to count an affirmation of propositional attitude as true or false, even given full knowledge of its circumstances and purposes" (1960: 218). The relativistic Rylean can countenance objective determinacy about whether a belief attribution is veridical or not: she simply must allow that an affirmation of propositional attitude is true or false *for a particular belief attributor*, and may not have the same truth value in relation to another attributor's model of belief.

primarily a property of behavior" (27; quoting Dewey 1925: 179), Quine went on to declare that "semantics is vitiated by a pernicious mentalism as long as we regard a man's semantics as somehow determinate in his mind beyond what might be implicit in his dispositions to overt behavior" (27). It would have been natural for Quine to proceed to defend the Rylean view that beliefs are nothing more than patterns of dispositions to behave in certain ways.[18]

But Quine's love of desert landscapes eclipsed his love of Dewey. In his heyday, Quine endorsed eliminativism about belief. While allowing that "I would not forswear daily use of intentional idioms" like belief, he insisted that "if we are limning the true and ultimate structure of reality, the canonical scheme for us is the austere scheme that knows no quotation but direct quotation and no propositional attitudes but only the physical constitution and behavior of organisms." Quine accepted "Brentano's thesis of the irreducibility of intentional idioms [as] of a piece with the thesis of the indeterminacy of translation." However, he went on to note that "one may accept Brentano's thesis either as showing the indispensibility of intentional idioms and the importance of an autonomous science of intention, or as showing the baselessness of intentional idioms and the emptiness of a science of intention." Middle-period Quine's "attitude, unlike Brentano's, is the second" (1960: 221). It is useful to talk about beliefs, but such talk is metaphysically spurious.

---

[18] However, insofar as a sparse Quinean Ryleanism would dispense with dispositions to think and feel, it would differ from Ryle's own view.

It would fall to Quine's successors to attempt to formulate realist, Rylean, but still unmistakably Quinean accounts of the nature of belief. Only after having assimilated two decades of work by his former students did Quine come around to the original version of Brentano's thesis. In his twilight years, Quine admitted, "thanks to Davidson's and [Dagfinn] Føllesdal's abetting" (1992: v), the importance of "irreducibly mental ways of grouping physical states and events." He concluded that talk of belief and the other attitudes engendered

> age-old strategies for predicting and explaining human action. They complement natural science in their incommensurable way, and are indispensable both to the social sciences and to our everyday dealings. Read Dennett and Davidson." (Quine 1992: 72–73)

Quine's command is my wish.

3.2. Dennett

Nearly thirty years before he came around to this Rylean way of thinking, Quine wrote a letter of recommendation that propelled Daniel Dennett from an undergraduate career at Harvard to graduate study with Ryle at Oxford (Dennett 2008a). Upon emerging from Oxford, Dennett began producing a body of work that demonstrated his enduring commitment to both Quine's preoccupation with indeterminacy and Ryle's

dispositionalism.

According to Dennett, interpreters have various 'strategies' or 'stances' they can adopt when predicting and explaining the behavior of objects in the world. Interpreters can adopt the physical stance, "and use [their] knowledge of the laws of physics to predict the outcome for any input" (Dennett 1987: 16). Or they can adopt the design stance, and predict that an object "will behave *as it is designed to behave* under various circumstances" (17). Finally, if the physical or design stances do not do the trick, interpreters can adopt the intentional stance, and treat an object "as an agent of sorts, with beliefs and desires and enough rationality to do what it ought to do given those beliefs and desires" (2009: 3).[19] Dennett explains that

> here is how [the intentional stance] works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in light of its beliefs" (1987: 17).

The intentional stance is the attitude interpreters take when they assume their subject

[19] Dennett now treats the intentional stance as "a subspecies of the design stance" (2009: 3), because his adaptationism leads him to consider animals who are intentional systems to have been designed by natural selection.

has (mostly) rational beliefs and (mostly) rational desires, and predict their subject's behavior accordingly.

For Dennett, "anything that is usefully and voluminously predictable from the intentional stance is, by definition, an intentional system" (2009: 1). My cats are intentional systems. I predict their behavior daily by assuming that they have (mostly) rational beliefs and desires, and find these predictions useful in determining when and how to pet, entertain, feed, and gently scold them. My mousepad, on the other hand, is not an intentional system. I *could* predict its behavior by assuming that it has a rational belief—the belief that it does not have anything better to do than keep on keeping on—as well as a rational desire—the desire to do nothing except continue to provide an amenable surface for my mouse. But such an adoption of the intentional stance is useless: I can predict my mousepad's (lack of) behavior perfectly well without pretending that it is a rational agent.

So far, Dennett's account is uncontroversial. Everyone agrees that it is often helpful to think about some objects (like people and cats and even chess-playing computers) as having rational beliefs and desires. But Dennett goes on to argue that there are no facts about what people believe other than facts about the most predictive ways of capturing how they represent the world from the intentional stance. On Dennett's view, believing and desiring—indeed having any intentional mental states—is nothing more than being usefully and voluminously predictable from the intentional stance.

As such, Dennett has often been taken to be an eliminativist (or, more precisely, an instrumentalist) about belief. The natural thought is that if "*what it is* to be a true believer is to be an *intentional* system, a system whose behavior is reliably and voluminously predictable via the intentional strategy" (1987: 15), then beliefs must not really exist. It must just be immensely useful to adopt the intentional stance and act as if people, cats and chess-playing computers had beliefs. But Dennett often calls himself a realist about belief (1987: 15; 1998: 98), and denies the charge of instrumentalism: "although the earliest definition of the intentional stance (Dennett 1971) suggested to many that it was merely an instrumentalist strategy, not a theory of real or genuine belief, this common misapprehension has been extensively discussed and rebutted in subsequent accounts" (2009: 14–15).[20] The rebuttals to which Dennett refers center on the notion that beliefs are elements of 'real patterns'.

Dennett introduces real patterns by reflecting on one of his favorite muses: Conway's Game of Life.[21] The Game of Life simulates a world comprising a grid of cells that are either on or off. The world progresses through time steps, and at each step the cells are governed by two simple laws of nature. If they are on, then they stay on if and only if two or three of their eight Moore neighbors are also on. If they are off, then they

---

[20] 'Misapprehension' is the wrong word, given that Dennett's earliest account explicitly "does not say that intentional systems *really* have beliefs or desires" (1971: 7). On my diachronic reading, Dennett began his career as a reluctant instrumentalist, and has since settled into the mild realism articulated in "Real Patterns" (1991a).

[21] Dennett uses the Game of Life to argue not only for the reality of belief, but also for the compatibility of determinism and free will (1984; 2013) and adaptationism in evolutionary biology (1995).

turn on if and only if three of their neighbors are on.[22] Striking visual patterns emerge from these simple rules: two-dimensional creatures composed of several cells seem to glide across the grid, only to be eaten by other, nastier creatures. From the rigorous perspective of the physical stance, of course, there is no way to discern multicellular entities or movement in the Game of Life; there are only cells that are on or off at different time-steps. But the visual patterns that we see as moving creatures are nevertheless real patterns arising directly from the cells adhering strictly to the programmed laws of physics.

The Game of Life is also remarkable for its computing power: in theory, its laws can be used to instantiate a Universal Turing machine, and "since the Universal Turing machine can compute any computable function, it can play chess – simply by mimicking the program of any chess-playing computer you like" (Dennett 1998: 109). Imagine programming the Game of Life to play chess against itself and then, after years of hard work, settling down to watch a match. In order to enjoy watching chess, you have to try to predict what moves the players are going to make. Of course, in our imagined scenario, you could adopt the physical stance. You have access to the vast number of inputs and the extremely simple laws of nature governing the Game, and it is therefore possible, in principle, for you to predict exactly what moves each artificially intelligent

---

[22] If you have not seen the Game of Life in action, I recommend getting out a piece of graph paper, marking some cells as on, and then executing time-steps by following the two laws. Alternately, the wikipedia page (http://en.wikipedia.org/wiki/Conway%27s_Game_of_Life) features gif illustrations, and this video (https://www.youtube.com/watch?v=xP5-iIeKXE8) provides an illustration of how the Game of Life can instantiate another version of itself.

player is going to make. But in practice, using the physical stance in this manner "would take eons" (1995: 235). A better idea is to adopt "the intentional stance towards the configuration [and] predict its future *as* a chess-player performing intentional actions – making chess moves and trying to achieve checkmate" (236). By deciding to treat the AIs as rational actors with (mostly) rational beliefs, you can predict what moves they are going to make in an efficient manner.

Dennett argues that just as we recognize real visual patterns of moving 2D creatures in the cells and time-steps of the Game of Life, there are other real patterns, composed of the same cells and time-steps, that we recognize as intentional systems. He writes that such "real but (potentially) noisy patterns abound in such a configuration of the Life world, there for the picking up if only we are lucky or clever enough to hit on the right perspective. They are not *visual* patterns but, one might say, *intellectual* patterns" (1998: 110). As understood from the intentional stance, the AIs instantiated in the chess-playing Game of Life are true believers. Likewise, as understood from the intentional stance, the people instantiated by human nervous systems are true believers. Dennett's interpretivism is not "merely an instrumental strategy" (2009: 15), because the intellectual patterns that constitute intentional systems with beliefs are not merely imposed on the world by the interpreter. Real patterns arise from the physical laws of the real world, just as they arise from the simple rules of the Game of Life. Dennett's theory of belief therefore may not be "Fodor's industrial-strength Realism (he writes it with a capital 'R')" (1998: 98), but it is nevertheless realism of "a mild and intermediate

sort." Dennett intends his discussion of real patterns to illustrate how mild realism is a "positively attractive position, and not just the desperate dodge of ontological responsibility it has sometimes been taken to be" (97).

Some philosophers (McLaughlin and O'Leary-Hawthorne 1994: 208; Child 1994: 50–51) have taken Dennett's commitment to real patterns to engender a rejection of interpretivism. Interpretivists think that to believe is to be interpretable as believing. However, according to Dennett, the real patterns that ground belief exist independently of the belief attribution practices of interpreters. Thus, although Dennett makes much of the power of the intentional stance, these philosophers think he must reject the idea that interpretation is actually constitutive of belief.

This reading is mistaken. Real patterns are not real patterns *of beliefs and desires* except insofar as they are subject to being interpreted in terms of "the underlying normative principle of intentional stance prediction – rationality" (1994: 523). Dennett has never given up on his earliest conviction that "a particular thing is an intentional system only in relation to the strategies of someone who is trying to explain and predict its behavior" (1971: 3–4). While the patterns that allow that thing to become an intentional system "are objective – they are *there* to be detected", they are nonetheless "not *out there* entirely independent of us, since they are patterns composed partly of our own 'subjective' reactions to what is out there; they are the patterns made to order for our narcissistic concerns" (1987: 39). In short, countless real patterns exist in the objective, perspective-independent world. But only some of those patterns are usefully

and voluminously detectable from the intentional stance. These particular patterns are rendered intentional—are "what Anscombe called the 'order which is there' in the rational coherence of a person's set of beliefs, desires, and intentions"—solely in virtue of being so detectable, and their elements are only thereby transformed from mere abstracta to (mostly rational) beliefs and (mostly rational) desires.[23]

Being a true believer, according to Dennett, is nothing more than giving rise to real patterns that are usefully recognizable from the intentional stance. But that still does not tell us what exactly beliefs themselves are. For that purpose, we need to invoke Dennett's distinction between "opinion" and belief: "between believing a certain sentence of your natural language is true, and having the sort of belief that the sentence might be used to express" (1998: 90). The former "are not just beliefs; these are linguistically infected states – only language users can have them. Opinions are essentially bets on the truth of sentences in a language that we understand" (89). Beliefs, on the other hand, involve not just attitudes towards propositions, but also, in true Rylean fashion, determinable dispositions to behave. The abstracta that the intentional stance transforms into beliefs are elements in real patterns (1998: 115–116). Those patterns are real because they are "indirect measurements of a reality diffused in the behavior dispositions of the brain (and body)" (1998: 114). When challenged by Frank Jackson to declare his exclusive allegiance, once and for all, to realism, behaviorism,

[23] Mölder (2010: 114–117) provides a nice discussion of this point, cast in the technical language of his own interpretivism.

functionalism, eliminativism, or instrumentalism, Dennett insists that his answer "mixes elements from all of these, and denies that there is any good reason to cleave to a less eclectic answer." He explains that

> *some* traditionally well-regarded mental states should be *eliminated*; in other words, only a reformed folk psychology stands in need of materialistic reduction. Now we must deal with the leftovers: what makes it true that people have the *real* mental states is facts about their *behavioral* dispositions and capacities, but these facts can be perspicuously and efficiently expressed only from the intentional stance, an *instrument* of prediction (and explanation). As for *functionalism*, in its defensible version, it is not really an alternative to behaviorism, but simply a reflection of the tight constraint behavioral capacities (as described from the intentional stance) place on internal states. So let me confirm Jackson's surmise that I am his behaviorist; I unhesitatingly endorse the claim that "necessarily, if two organisms are behaviorally exactly alike, they are psychologically exactly alike." (Dennett 1993: 923, quoting Jackson 1993: 4)

Thus, for Dennett, beliefs are abstract elements of intentional systems that perspicuously and efficiently express (by indirectly measuring) real behavioral dispositions.

So much for Dennett's Rylean streak. Time to bring Quine into the fold.

The intentional stance can backfire. Sometimes the people, animals and cartoon

characters we expect to be intentional systems act irrationally. But Dennett is not overly concerned with cases like these. Usually, positing false beliefs and perverse desires is the most efficient way to restore the coherence of the presumption of basic rationality. In other cases, the design stance can kick in and provide a fine explanation for behaviors that seem bizarre from the intentional stance.

A variety of case that is more important to Dennett's thinking involves the intentional stance being prodigiously efficacious. Dennett holds that the intentional stance can be so fecund that it generates multiple interpretations of an intentional system that are equally good at predicting behavior. What is more, Dennett holds that sometimes these interpretations can substantively conflict when it comes to particular belief attributions and particular predictions of behavior: "there could be two interpretation schemes that were reliable and compact predictors over the long run, but which nevertheless disagreed on crucial cases" (1998: 117). When this is the case, Dennett embraces his Quinean side and claims that what the intentional system really believes is fully indeterminate. There is no deeper, intersubjectively discernable fact of the matter—no physical or functional state of belief in the brain, for instance—to appeal to.

Dennett asks us to consider the case of Brown and Jones, concerned friends who disagree about whether "Ella will decide to kill herself" (1998: 115). Brown predicts that she will; Jones predicts that she will not. Perhaps Brown arrives at his prediction by imputing to Ella the rational beliefs that life is full of pain and that in death there is no

pain. Perhaps Jones arrives at her prediction by imputing to Ella the rational beliefs that it would hurt her family to commit suicide and that some things make life worth living. These beliefs are just a couple of points in the vast holistic patterns of beliefs arising from the way Ella leads her life. Brown and Jones "agree on the general shape of [Ella's] collection of beliefs (and desires, etc.), but because of their different idealizations of the pattern, they do not agree point-for-point" (115). In this case, the points of disagreement lead to importantly different predictions about Ella's behavior. But Dennett claims that even "if Ella does kill herself … Brown [is] not necessarily … shown to have clearly had the better intentional interpretation" (117). One bad prediction does not condemn an interpretive strategy. If, over the long run, the pattern Jones interprets Ella as embodying is just as usefully and voluminously predictive as the pattern that Brown interprets Ella as embodying, then both patterns are real. Dennett digs in his heels alongside Quine and claims that "no deeper fact of the matter could establish that one was a description of the individual's *real* beliefs and the other not" (118). Thus, for Dennett, there is no determinate way of deciding whether Ella believes that life was or was not worth living. There is no good reason to think that rooting around in Ella's brain could reveal a state of neural activity that would determinately tip the balance in either Jones or Brown's favor.[24]

---

[24] Dennett insists that while the conceivability of full intersubjective indeterminacy is "metaphysically important" for determining the nature of belief, its results are "practically negligible" (1987: 29). I agree with Dennett about the former, but disagree about the latter. As I will argue in Chapters 7 and 8, Ryleans should hold that intersubjective indeterminacy in belief attribution can sometimes lead to practically significant consequences.

3.3. Davidson

While Donald Davidson agreed with Dennett that to believe is a matter of being a suitable object of interpretation, their interpretivisms diverge in key respects. The main lesson Dennett takes from Quine is that there is not always a determinate, intersubjective fact of the matter about what people believe. Davidson, on the other hand, took indeterminacy to be a red herring, and instead followed Quine mainly by insisting that belief is grounded in social practices that involve the interpretation of each other's speech.

Most treatments of Davidson's interpretivism (Fodor & Lepore 1992: 59–104; Heil 2003: 131–154; Mölder 2010: 85–102) focus on his writings from the 1970s, and therefore begin (and usually end) with an analysis of how Davidson's account of radical interpretation augmented Quine's notion of radical translation. I will take a different tack, drawing on Davidson's early writings only insofar as they inform his mature views (developed through the end of the 20th century) on the nature of belief. Davidson's understanding of radical interpretation, especially as elaborated by the metaphor of triangulation, is crucial to his understanding of belief. But in the '70s, Davidson invoked radical interpretation not to provide a theory of belief, but primarily to argue that Alfred Tarski's (1944) theory of truth, wielded by a radical interpreter, could be used to establish an empirical theory of meaning without succumbing to full indeterminacy. In the '80s and '90s, Davidson committed to articulating a theory of belief (and closely related theories of meaning and objectivity). Davidson's overarching philosophical

project was to explain how, the inscrutability of reference notwithstanding, we are able to use language to learn objective truths about the world (including beliefs).

Whereas Dennett's interpretivism is founded on the principle that full Quinean indeterminacy is inevitable, Davidson's was founded on the principle that it is inconceivable.[25] Davidson acknowledged the inscrutability of reference. But, following Quine, he denied that this sort of weak indeterminacy is problematic. To show why, Davidson drew an analogy between the indeterminacy of translation and the use of different measurement systems.

> It should not bother us that the Frenchman and I use different utterances to characterize the same state of Paul's mind: this is like measuring weight in carats or ounces: different sets of numbers do the same work … When we use numbers to keep track of the relations among weights and lengths and temperatures, we are not apt to respond to the fact that different sets of numbers do as well as others in keeping track of all that is relevant empirically by complaining that weights or lengths or temperatures are not 'real'. We know that there is no contradiction between saying that the temperature of the air is 32° Fahrenheit and saying that it is 0° Celsius; there is nothing in this 'relativism' to show that the properties being measured are not real. (1989: 64–65)

[25] Again, I challenge this principle in Chapters 7 and 8.

Davidson argued that it does not matter that it is indeterminate whether 'Gavagai' means 'rabbit' or 'rabbit time slice' because the radical translator knows, either way, that 'Gavagai' is used to track the same empirical phenomenon that she tracks using the word 'rabbit'. Moreover, Davidson departed from Quine and Dennett by insisting that this innocuous form of indeterminacy is the only way in which the meanings of our words, as well as the contents of our beliefs, can be indeterminate.

For Davidson, the idea that meanings and contents can be fully indeterminate invokes the untenable idea of "conceptual schemes", that is, potentially incommensurable "ways of organizing experience" (1974: 183). Davidson argued that Quine was blinded by "the dogma of a dualism of scheme and reality" (198), according to which there are one or more schemes that foist idiosyncratic interpretations onto a scheme-neutral, uninterpreted reality. Davidson rejected this dogma in favor of the view that every viable interpretation involves coming to grips with the world as it actually—objectively, intersubjectively, and subjectively—is. Like Quine (1951: 39–43) and Dennett (1998: 89), Davidson accepted the holism of the mental, writing that "there are … no beliefs without many related beliefs, no beliefs without desires, no desires without beliefs, no intentions without both beliefs and desires" (1997: 126). But he rejected the Quinean and Dennettian idea that interpreters could latch onto holistic webs of belief that were fundamentally at odds with one another.

On Davidson's view, it is inconceivable that there be competing interpretations that result in an indeterminacy more problematic than the indeterminacy between

whether the temperature outside is 32°F or 0°C. Just as there is always a correct answer to the question whether or not the temperature is below the freezing point of water, it is always the case that "there is a correct answer to the question whether or not someone has a certain attitude" (1997: 82). Thus, Davidson would reject Dennett's thought experiment concerning Brown and Jones's conflicting interpretations of Ella as inconceivable.[26] For Davidson, although Ella's set of beliefs can be picked out in different ways, there must be some fact of the matter that precludes full indeterminacy. As he put it, "Dennett has urged that the answer to the [question whether there are objective grounds for choosing among conflicting hypotheses] is that there are no such ground; but I do not think he has given any reason to accept this answer" (82). As it happens, Davidson's lifelong philosophical project can be seen as an attempt to understand how such objective grounds are possible.

How, then, did Davidson think Brown and Jones could get at Ella's one true and determinate set of beliefs and desires? In short: the same way anybody gets at any truths. One major upshot of Davidson's philosophy was that "if our judgements of the propositional attitudes of others are not objective, no judgements are, and the concept of objectivity has no application" (1997: 84). For Davidson, objectivity springs from intersubjectivity—from agreement between interpreters—and intersubjectivity springs from an understanding of each other as minded subjects. Davidson held the objectivity

---

[26] Davidson also criticizes Dennett's emphasis on "the issue of prediction [as] something of a red herring" (1997: 81). He thinks it is a red herring in part because his anomalous monism, discussed below, precludes lawlike inferences from mental states to behavior.

of interpretations of others as believers and the objectivity of knowledge of the external to arise hand-in-hand because

> the ultimate source (not ground) of objectivity is, in my opinion, intersubjectivity. If we are not in communication with others, there could be nothing on which to base the *idea* of being wrong, or, therefore, of being right, either in what we say or in what we think. (1997: 83)

In order to have beliefs, people must have a sense of objectivity, and "our sense of objectivity is the consequence of [a] sort of triangulation … that requires two creatures [and] an object" (1982: 105). By communicating with other creatures and thereby triangulating on objects (including our own beliefs and the beliefs of our interlocutors), people are able to establish that those objects really, intersubjectively and therefore objectively exist (rather than only seeming to exist from a particular belief attributor's first-person point of view.) How can Brown and Jones come to know what Ella really believes? By talking to each other, coming to understand each other's points of view, and thereby triangulating on the objective truth.

Davidson's first step towards this irreducibly social account of belief was to promote Quine's radical translator—concerned only with "the syntactic notion of translation"—to a radical interpreter—concerned with "the semantic notion of truth" (1983: 148). Unlike the radical translator, the radical interpreter has recourse to two

important interpretive devices: the notion of 'holding true' (an ancestor of Quine's 'holding true come what may') and the principle of charity as applicable to the semantic content of beliefs. The radical interpreter needs this extra interpretive power because her task is extra difficult. Whereas Quine's radical translator focused on the comparatively straightforward job of translating utterances, the radical interpreter must also figure out both the truth conditions for a speaker's utterances "and a theory of his beliefs" (148). What is more, Davidson held "that interpreting an agent's intentions, his beliefs, and his words are parts of a single project, no part of which can be assumed to be complete before the rest is" (1973: 127). The radical interpreter cannot use what an agent means by his sentences to figure out what he believes (or vice versa), because she cannot know what he means (or believes) independently of knowing what he believes (or means).

Nevertheless, Davidson did allow the radical interpreter access to one kind of attitude from the start: "the attitude of holding a sentence true, of accepting it as true." He wrote that "an interpreter may plausibly be taken to be able to identify [the attitude of holding true] before he can interpret, since he may know that a person intend to express a truth in uttering a sentence without having any idea *what* truth" (135). By using attributions of holding true as a kind of Archimedean point, the radical interpreter can charitably assume that the agent has "a set of beliefs largely consistent and true by [the interpreter's] own standards" (137)—much like Dennett's interpreter who assumes rationality from the intentional stance. Davidson held that a radical interpreter—"torn between the need to make sense of a speaker's words and the need to make sense of the

pattern of his beliefs" (1968: 101)—must adopt this principle of charity lest she "have no reason to count that creature as rational, as having beliefs, or as saying anything" at all (1973: 137).

However, leveraging these interpretive devices can only get the radical interpreter so far. Over the course of his career, Davidson increasingly came to insist that in order to complete the project of interpretation the tables have to turn; the interpreter has to become the interpretee, and both radical interpreters (the old and the new) have to reflect on their own beliefs, the other's beliefs, and the disparate objects in the world that both sets of belief are about. Davidson identified "three basic problems: how a mind can know the world of nature, how it is possible for one mind to know another, and how it is possible to know the contents of our own minds" (1991: 208). He argued that these problems are solved via triangulation, all at once or not at all, "for the triangulation which is essential to thought requires that those in communication recognize that they occupy positions in a shared world" (213). In particular, triangulation "forces us to the idea of an objective, public truth" (1975: 170).

It does not take long, when two people start talking, for it to become starkly obvious that sometimes one of the interlocutors must be wrong. Consider Brown and Jones again. As soon as they start arguing, it becomes readily apparent to both that there is a determinate truth about what Ella believes (even if the truth is that Ella has a Schwitzgebelian in-between belief). What the determinate truth *is* may remain indefinitely unclear. However, for Davidson, the very fact that Brown and Jones can

make sense of their disagreement shows that Ella must have some objective belief or other about which Brown and Jones disagree.

Davidson's most famous contribution to the philosophy of mind was his anomalous monism: the view that token mental states and events are identical to token physical states and events, but there are no strict laws that relate mental events to physical events (or mental events to one another). Davidson introduced anomalous monism in a 1970 paper titled "Mental Events", and focused his argumentative attention on the eponymous events—"such as perceivings, rememberings, decisions, and actions"—because his aim was to defend the "*anomaly* (failure to fall under a law)" of the mental (207). Only events fall (or fail to fall) directly under laws. Because of his anomalous monism and attendant focus on events, scholars often take Davidson's view to have been that belief is a kind of mental event, and that token beliefs are therefore identical to token physical events (Mölder 2010: 86; Yalowitz 2012). This interpretation is mistaken. While Davidson refers to *changes* of belief as mental events (208), he never refers to beliefs themselves as events.

Closer inspection reveals that Davidson's actual view was that beliefs were anomalously identical not to physical events—such as patterns of neural activity—but to the physical bases for sets of behavioral dispositions. Davidson held that "anomalous monism makes sense of the claim that attitudes are dispositions to behave in certain ways, which are in turn physiological states, which finally are physical states, as well as the claim that intentional descriptions are not reducible to behavioral or physical

descriptions" (1997: 72). On Davidson's account, beliefs are determinable "dispositions that manifest themselves in various ways, over a span of time" (1984: 4). According to his anomalous monism, these patterns of dispositions are irreducibly identical to whatever physical processes underlie them.

Other scholars have inferred that Davidson was an interpretivist about belief only insofar as he was an interpretivist about everything (Schroeder 2003). After all, Davidson held that we obtain objective knowledge of our world (including knowledge of believers) only insofar as we become radical interpreters who have triangulated on the world with other radical interpreters. But two of Davidson's commitments—to the anomaly of the mental and the principle of charity—show that this could not have been his view. Although we come to know physical states and events exclusively via triangulation, those states and events are subject to the laws of physics whether or not we know about them. The lightning strikes the tree and causes it to crash to the ground, creating sonic reverberations, whether or not there are radical interpreters in the forest to hear the commotion. Belief, on the other hand, emerges in the world only with the emergence of interpretation; "there is no propositional thought without communication" (1991: 213). An animal who is not (yet) a believer or radical interpreter may be disposed to behave in ways consonant with taking the world to be a certain way. But that pattern of dispositions is not a belief unless (or until) the animal, through linguistic interaction with another, reifies it as such "by deciding [that the attribution of belief], from his point of view, maximizes intelligibility" (215). According to Davidsonian interpretivism,

mental states are irreducibly real, but metaphysically emerge only in the intersubjective context of triangulation. To believe is to be objectively and veridically interpretable as believing.

Elaborations and defenses of the pioneering work of Dennett (Wilkerson 1997; Zawidzki 2013, 2015) and Davidson (Gauker 1988, 1994, 2003; Child 1994; Verheggen 2007, 2016) notwithstanding, interpretivist theories of belief have not been as popular as their dispositionalist cousins. Philosophers have produced a great deal of work on *how* we interpret others as having beliefs (Bogdan 1997, 2000, 2008; Maibom 2003; Nichols & Stich 2003; Godfrey-Smith 2005; Goldman 2006; Hutto 2008; Lurz 2011; Andrews 2012), and some philosophers have considered interpretivism about some subset of beliefs (Haugeland 2002; Kriegel 2011), but philosophers have by and large shied away from all-out endorsements of the theory. Bruno Mölder is a recent exception.

3.4. Mölder

In his 2010 book *Mind Ascribed*, Mölder provides the most systematic interpretivism yet on offer, which he terms 'the ascription theory'. The systematicity of the ascription theory derives from a certain simplicity. Mölder holds that "the study of mind is the study of the concept of mind" (132). In particular, Mölder defends the view that all there is to having a mind is "the possession of mental states and the possession of mental contents" (2), and that all there is to possessing mental states is being

interpretable as possessing those mental states in the right sort of way.[27] Mölder terms the right sort of interpretation a 'canonical ascription'. In short, as with all other mental phenomena, "to have a belief that *p* is to have it canonically ascribable" (171).

Mölder gives two conditions for canonicality: "coherence and revisability" (14). A belief is canonical just in case its hypothetical ascription would be maximally coherent with all of the evidence that is "in principle accessible to ordinary folk" (174) and, "if the ascription were in fact made, it would not *require revision*" (175). Mölder imposes the first, accessibility condition in place of the rationality constraint (or principle of charity) imposed by both Dennett and Davidson. (I discuss the second, revisability condition shortly.) Mölder agrees with critics (Stich 1981, 1990; Nisbett et al. 1983; Goldman 1986) of Dennett and Davidson that "people are not mostly rational" (118), but is "inclined to think that the quarrel over rationality is a red herring in discussions about interpretivism" (119).[28] Mölder maintains that

> we can make progress by shifting the focus from the rationality issue to what really matters for interpretivism. What really matters is whether a subject's mental states and contents can be inferred or constructed on the basis of the available evidence (2010: 119).

---

[27] Mölder uses "mental states" to include "events, and processes" (2010: 148).
[28] I agree.

He lists four "ascription sources", or kinds of evidence that constrain canonical ascriptions of belief: "*other mental states with contents that are ascribable*", "a subject's *behavior*", "a subject's *environment*", and "a subject's *personal background*, which involves facts about the subject's personal history and dispositions, their language community and idiosyncratic language understanding" (161–162). The first condition of canonicality is fulfilled if the hypothetical ascription makes as much sense of all the evidence from these sources as could be expected, given that the interpreter is just some guy—an actual human being rather than a hypothetical omniscient being.

Mölder also gives two conditions for ascribability, concerning "facts about the subject and its environment" and "facts about canonical ascribers" respectively. A belief is ascribable "in virtue of facts that are collected in ascription sources and facts about how ordinary ascribers interpret subjects" (178). Ordinary ascribers interpret subjects "in terms of folk psychology" (160). If the evidence about an object of interpretation is such that ordinary folks considering that evidence might ascribe belief to that object, then belief is ascribable to that object. Putting the ascribability conditions and first canonicality condition together, Mölder's thesis is that having a belief is primarily a matter of having other characteristics that folks might wield as evidence when ascribing the belief to you, so long as the ascribable belief coheres with those characteristics.

Mölder follows a very Dennettian route to both realism and interpretivism. For Mölder as for Dennett, beliefs are grounded in real physical characteristics, but only count as beliefs insofar as there are folk psychological practices in place that treat them

as such. Unlike both Dennett and Davidson, however, Mölder takes belief to exist relative to canonical—community standard—attributions, rather than (objectively) normatively ideal attributions.[29]

Mölder adds the second canonicality condition—that, in context, the ascription must not require revision—in order to avoid the sort of full indeterminacy that Dennett embraces. The second condition allows Mölder to claim that "*incompatible* mental profiles are not canonically ascribable" (183) because

> the fact that the canonical ascriber is faced with two different or incompatible ascriptions is a good reason for seeking further revisions. In that case, substantially different ascriptions are not both canonical. (2010: 184)

If Brown and Jones both have complete evidence but still disagree about whether Ella believes that life is worth living, then at least one of their interpretations requires revision, and thus is not canonical.

The three (or four, if we count the octogenarian Quine) interpretivists discussed in this section have distinct projects and distinct commitments. Dennett takes interpretations to be apt if they are maximally useful for prediction, Davidson if they are true (given the objectivity that arises from intersubjectivity), and Mölder if they are

[29] While I applaud Mölder's rejection of ideal interpreters, I argue against his invocation of community standards in Chapter 7.

canonical. Nevertheless, the theories of belief they devise share a common, Quinean and Rylean core. For Dennett, Davidson and Mölder, to believe is nothing more than to be aptly interpretable as believing.

4. Ryleanisms

I have given an overview of six theories of belief: three dispositionalisms and three interpretivisms. We are now well positioned to see that the three dispositionalisms reviewed are also interpretivisms and that the three interpretivisms reviewed are also dispositionalisms. This convergence is no coincidence; nor is it the result of my having cherry-picked philosophers. Instead, it is the result of dispositionalists and interpretivists having a common enemy: the standard view that the beliefs people attribute to each other in everyday life—if they exist—are theoretically posited cogs playing functional roles in cognitive machines. Following Ryle (1949: 18–24), I will henceforth refer to the standard view as the 'paramechanical view of belief', and advocates of this view as 'paramechanists'.

In this section, I will argue that all dispositionalisms that seek to rival paramechanical versions of the mind-brain identity theory, functionalism, and eliminativism are *ipso facto* interpretivisms, and that all interpretivisms sharing that goal are *ipso facto* dispositionalisms. Insofar as they promise alternatives to the dogma that the beliefs lay people attribute are theoretical entities, there are no pure dispositionalisms or pure interpretivisms; there are only a variety of interpretivist

dispositionalisms, or dispositionalist interpretivisms, or, as I will call them from here on out, 'Ryleanisms about belief'.

From the trio of Ryle, Schwitzgebel, and Baker, we have gleaned what all dispositionalisms have in common. From Dennett, Davidson, and Mölder we have gleaned what all interpretivisms have in common. By glossing over individual philosophers' idiosyncratic commitments, we arrived at the following minimal theories of belief.

Minimal dispositionalism: To believe is to have an appropriate pattern of dispositions.
Minimal interpretivism: To believe is to be aptly interpretable as believing.

Minimal dispositionalism and minimal interpretivism are both complete theories of belief. For the minimal dispositionalist, having an appropriate pattern of dispositions is necessary and sufficient for belief; for the minimal interpretivist, being aptly interpretable as believing is necessary and sufficient for belief. However, each minimal theory also contains a weasel word. A lot can be packed into the *appropriateness* of a pattern of dispositions, or the *aptness* of an interpretation. My arguments in this section hinge on pointing out how dispositionalists (sometimes surreptitiously) squeeze interpretivism into their accounts of what makes it appropriate to equate a pattern of dispositions with a belief, and interpretivists squeeze dispositionalism into their account of what makes an interpretation apt. Dispositionalism and interpretivism are two sides

of the same coin. Each of the philosophers discussed in this chapter defends some version of the minimal Rylean view that to believe is to possess a pattern of dispositions that an interpretive scheme designates as constitutive of belief.

Dispositionalists writing after Dennett and Davidson have been quick to distance themselves from the interpretivisms on offer. Baker (1995: 237, 189) rejects Dennett's instrumentalism and Davidson's insistence on the co-dependence of belief and language. Schwitzgebel (2012: 19) rejects interpretivism's reputed overreliance on observable behavior. But minimal interpretivism has no commitments to anti-realism, the dependence of belief on language, or reliance on observable behavior alone. (Indeed, all three of our interpretivists reject the first, Dennett and Mölder reject the second, and Davidson and Mölder reject the third.) As we have seen, Baker also stresses that her dispositionalism is recognition-independent. But recognition-independence, unlike mind-independence, is built into minimal interpretivism, which makes only interpretability (rather than occurrent interpretation) a prerequisite for belief. At least in their minimal forms, dispositionalism is perfectly compatible with interpretivism.

Indeed, the two theories are more than just compatible. Despite their reservations, each of our dispositionalists is committed to minimal interpretivism.

For dispositionalists, to believe is to have an appropriate pattern of dispositions. There must be some criterion for what makes a pattern of dispositions the appropriate pattern to constitute belief. In other words, there must be something that systematizes the variety of dispositions to behave, think and feel as if the world is a certain way. The

skater who believes that the ice is thin is disposed to skirt the ice, to feel uneasy, to say "be careful out there!", to suggest his kids stay inside and drink hot chocolate instead of going skating, and so forth. The question at hand is what makes all of these disparate dispositions elements of the same belief.[30] Why *this* particular pattern of dispositions rather than some other pattern?

So, what could make it such that *this* rather than *that* pattern of dispositions is appropriately identified with a belief? One plausible criterion is that there is a single—though perhaps distributed—neural state (or functional role) underlying all of the dispositions in the appropriate pattern. But if the dispositionalist accepts this criterion, then dispositionalism would not be a true rival to paramechanical theories. The theory that 'to believe is to have a pattern of dispositions that share a neural basis (or functional role)' collapses into the theory that 'to believe is to have a certain neural state (or have a cognitive state that functions in a certain way)'. Both the mind-brain identity theory and functionalism, in their paramechanical forms, dictate beliefs to be theoretically posited cogs in cognitive systems that produce behaviors, thoughts, and feelings. The dispositionalisms reviewed in this chapter are supposed to embody the alternative

---

[30] How dispositions are appropriately systematized can vary from context to context. For example, the dispositions that realize a given belief vary from believer to believer based on their other beliefs. If Roger believes that it is best to sleep in whereas Samantha believes that the early bird gets the worm, then the dispositions that are appropriately identified with their other beliefs may vary, sometimes to the point of contradiction. Roger and Samantha might both believe that the sun will rise tomorrow morning, for instance. But in Samantha's case, this belief will be identified with a pattern of dispositions that includes her disposition to open the blinds tonight. In Roger's case, the belief may be identified with a pattern that includes his disposition to close the blinds tonight.

suggestion that beliefs are ways of living that explain why believers behave, think and feel as they do without invoking theoretically posited producing causes.

Interpretivism provides the alternative criterion of appropriateness which allows dispositionalism to be a true rival to paramechanical views. An interpretive scheme designates a pattern of dispositions as constitutive of belief. Indeed, each of our three dispositionalists adopts some version of this interpretivist criterion of appropriateness.[31]

Here, then, is my argument that all dispositionalisms that rival paramechanical views must invoke minimal interpretivism.

P1. Per dispositionalism, to believe is to have an appropriate pattern of dispositions.

P2. There must be some criterion for what makes a pattern of dispositions appropriate.

P3. Candidate criteria are (a) that the relevant dispositions are grounded in a physical or functional state of belief or (b) that they are grouped by an interpretive scheme.

P4. If (a), dispositionalism collapses into a paramechanical version of the mind-brain identity theory or functionalism.

---

[31] Ryle relied on the logic of belief thought and talk to serve as the interpretive scheme that unveils which determinable dispositions are identical to which beliefs. Baker calls upon our shared commonsense conception. Schwitzgebel, unhappy with the vague pronouncements of both of his predecessors on this score, identifies interpretive schemes with dispositional stereotypes.

P5 If (b), dispositionalism involves interpretivism.

C. If dispositionalism is an alternative paramechanical views, then it involves interpretivism.

Dispositionalism is a rival to paramechanical versions of the mind-brain identity theory and functionalism if and only if it can explain the nature of belief without reference to physical or functional states underlying belief. However, this mark of heterodoxy does not entail that dispositionalism must preclude the theoretical possibility that unified physical or functional states underlie all of the dispositions that realize particular beliefs. To retain its anti-paramechanical credentials, dispositionalism only need show how we are warranted in asserting the existence of beliefs, qua patterns of dispositions, whether or not such neural or cognitive categorical bases exist.[32]

My argument for interpretivists' surreptitious commitment to dispositionalism closely parallels my argument for dispositionalists' commitment to interpretivism. According to minimal interpretivism, to believe is to be aptly interpretable as believing. There must be some criterion for what makes an interpretation apt. For Dennett, that criterion is being useful for the purpose of predicting behavior (which entails tracking a real pattern), for Davidson the criterion is being intersubjectively determined to be objectively true, and for Mölder the criterion is being canonical.

But what makes an interpretation useful, true, or canonical? The answer must

[32] I back up this assertion in Chapter 5.

rely on something about the believer being interpreted. In particular, there have to be relevant facts about the believer that hold (or fail to hold) prior to any interpretation. Otherwise, it would be impossible to make sense of interpretive error. An interpreter can be wrong to attribute the belief that the ice is thin to the skater. If the skater does not believe she is on thin ice, then the interpretation has failed. Depending on who you ask, the interpreter has failed to track a real pattern, or to triangulate thoroughly, or to make her interpretation cohere with the ascription sources. In any case, there must be something about the target of interpretation that makes the interpretation a failure. Ever the Rylean, Dennett makes clear that the relevant facts about the target of interpretation are "behavior dispositions of the brain (and body)", indirectly measured from the intentional stance (Dennett 1998: 114). Davidson echoes this dispositionalist thought: "though there is first person authority with respect to beliefs and other propositional attitudes, error is possible; this follows from the fact that the attitudes are dispositions that manifest themselves in various ways, over a span of time" (Davidson 1984: 4).

Mölder acknowledges his admiration for Baker's Practical Realism, and argues that "Baker's [dispositionalism] can be made compatible with interpretivism" (2010: 157). Nevertheless, Mölder stresses that "interpretation is underdetermined by a subject's behavior" (123), and that his "approach involves expanding the sources of evidence that are relevant in interpretation" (122). As such, Mölder takes himself to reject the idea (which he associates with Dennett and Davidson as well as Baker) that behavioral dispositions alone can serve to make interpretations apt or inapt. Mölder's

first canonicality condition provides four ascription sources: not only the subject's behavior, but also other canonically ascribable mental states, the subject's environment, and the subject's personal background. But it turns out that each of these ascription sources is important strictly as an epistemic tool for getting at a pattern of dispositions. Beliefs might be canonically ascribable because they cohere with other mental states, but those mental states in turn (or, anyway, some mental states down the line) must be canonically ascribable based on the subject's dispositions. Meanwhile, "the environmental *stimuli* that have an effect on the subject" are an ascription source because "in conjunction with other sources [they are] important in yielding the right sort of mental profile" (162). In other words, the subject's environment matters because it constrains which dispositions the subject can have. The subject only has the disposition to kick the can in environments in which the can is on the scene. Finally, the subject's personal background is "a ragbag" (163) collection of facts about the subject (her history, her language, and so on) that all serve to show that actual behavior is not a proper basis for belief; only dispositions (including dispositions to behave) will do. In the end, then, Mölder agrees with Dennett and Davidson that dispositions (and dispositions alone) serve to make an interpretation apt, while reasonably stressing that the aspiring ascriber of belief must consider a variety of kinds of evidence in order to determine which dispositions her subject possesses.

Without further ado, here is my argument that interpretivisms that rival paramechanical views must involve minimal dispositionalism:

P1. Per interpretivism, to believe is to be aptly interpretable as believing.

P2. There must be some criterion for what makes an interpretation apt.

P3. In order for there to be such a criterion, there must be some fact that holds about believers independently of our taking them to believe.

P4. Reasonable candidate facts are (a) that the believer has a physical or functional state of belief or (b) that the believer has an appropriate pattern of dispositions.[33]

P5. If (a), interpretivism collapses into paramechanical versions of the mind-brain identity theory or functionalism.

P6. If (b), interpretivism involves dispositionalism.

C. If interpretivism is an alternative to paramechanical views, then it involves dispositionalism.

---

[33] There are other candidate facts, for example (c) that the believer has an appropriate set of occurrently manifest behaviors, (d) that the believer has an appropriate set of past behaviors and experiences, or (e) that the believer belongs to a particular ethnic, political, or religious group. But I propose that all candidate facts, other than (a) and (b), have absurd consequences. For example, (c) would rule out the interpretation that I believe I have hands whenever none of my occurrent behaviors exhibit that belief, (d) would rule out the interpretation that I've forgotten where I put my keys, and (e) would rule out the possibility of prodigal sons, mavericks, and heretics. One might think that (c) and (d) combined make for a good basis for the aptness of an interpretation, but the combined candidates still fail to account for behavior in new environments—see Mölder's argument (2010: 123), drawing on Byrne (1998), that "interpretation is underdetermined by a subject's behavior." In order to expand the account such that (c) and (d) do account for new environments, the emphasis on actual behavior would have to be given up and dispositions invoked: the (c–d) theorist would have to say what the believer *would* do in the new situation—for example, what the skater who believes the ice is thin *would* do if (counterfactually) stranded on the middle of the lake—and thereby invoke dispositions.

Again, an alternative to bringing dispositionalism on board is to admit that every belief is backed up by a neural state or functional role. According to that alternative, an interpretation would be apt just in case the believer actually has the relevant brain state, or has a cognitive state the plays the relevant functional role. But admitting physical states or functional roles as the categorical bases for beliefs would lead interpretivism to collapse into paramechanical versions of the mind-brain identity theory or functionalism respectively (augmented by the facile claim that we have the capacity to interpret each other as believing). Interpretivism rivals paramechanical views insofar as it does not invoke inner states that produce behavior. Of course, interpretivists can admit that the states posited by realist adherents to paramechanical views *might* exist. Interpretivists are even free to wager that the relevant brain states or functional roles *do* exist: that humans have a neurally realized Language of Thought comprising syntactic representations underlying all beliefs, for example. But interpretivism, mixed with dispositionalism, is an interesting heterodox theory of belief insofar as it purports to show that (and how) beliefs exist *whether or not* cognitive scientists will one day discover brain states or functional roles that constitute the categorical bases for all believing. Any interpretivism that wants to achieve this aim has to bring dispositionalism on board in order to account for what makes interpretations (in)apt.

All in all, dispositionalism and interpretivism are not two distinct types of anti-paramechanical theory of belief. There is only one going heterodox realism in the philosophy of mind: Ryleanism, versions of which stress dispositions over

interpretations or vice versa. In its minimal form, Ryleanism is the theory that beliefs are patterns of dispositions designated as beliefs by an interpretive scheme.

Minimal Ryleanism: To believe is to have a pattern of dispositions that an interpretive scheme designates as constitutive of belief.

Over the course of the next three chapters, I will clarify what is at stake in proposing Ryleanism as a theory of belief, and defend minimal Ryleanism against paramechanical versions of the mind-brain identity theory, functionalism, and eliminativism. In Part II, I will introduce a novel version of Ryleanism—comprising new characterizations of the relevant interpretive schemes and the patterns they pick out—and argue for its merits over the other Ryleanisms reviewed in this chapter.

Chapter 2: Varieties of (theory of) belief

1. Introduction

In this chapter, I will canvas varieties of belief. This canvassing will build up to a crucial distinction between the attitudes of belief that people attribute to each other in ordinary life and the cognitive states of belief posited by (some) cognitive scientists. I will argue that Ryleanism about belief diverges most significantly from standard theories of belief insofar as it refuses to conflate attitudes of belief with cognitive states of belief, and conclude by noting that Ryleanisms are therefore compatible with the mind-brain identity theory, various functionalisms, and eliminativism about cognitive states of belief. For example, purported deep divides between dispositionalists and representationalists (Schwitzgebel 2002; Quilty-Dunn & Mandelbaum 2017) are little more than smoke and mirrors: the former offer an account of attitudes and the latter offer an account of cognitive states.

Most humans routinely attribute beliefs for (at least) the purposes of predicting and explaining anomalous behavior, managing impressions, making ethical and aesthetic evaluations, regulating the behavior of others, figuring out how to behave themselves, and gathering evidence about what is likely to be true. These beliefs range from simple judgments—for instance a cat's perceptual belief that there is kibble in her bowl—to judgments concerning linguistically expressed propositions—a physicist's belief that $e=mc^2$—to personal philosophies—an unwitting Kantian's belief that lying is

always wrong—to articles of faith—a zealot's belief in God, or a patriot's belief in her country. I will call all of these states of belief that humans frequently attribute to each other in everyday life '*attitudes* of belief'. In Ian Hacking's terminology, attitudes of belief are paradigmatic 'human kinds': "digests of what matters in intimacy" (1995: 354). There are distinctive metaphysical (and methodological) quandaries about how to characterize attitudes of belief, as well as about whether or not attitudes of belief exist.

Meanwhile, some cognitive scientists attribute beliefs to humans and other rational animals for the purpose of providing a scientific explanation of how those animals cognize objects, and then proceed to behave appropriately in their environments. I will call the states of belief posited by cognitive scientists '*cognitive states* of belief'. If "cognition is what supports intelligent behavior" (Hatfield 2014: 363, reflecting on the legacy of Haugeland 1981), then cognitive states of belief are postulated inner producers of intelligent behavior. If attitudes of belief are paradigmatic human kinds, then cognitive states of belief are paradigmatic natural kinds, displayed in high relief when sophisticated cognitive systems are carved at the joints. Again, there are distinctive metaphysical (and methodological) quandaries about how to characterize cognitive states of belief, as well as about whether or not cognitive states of belief exists.

In Chapter 1, I insinuated that there is a deep disagreement between philosophers in the Rylean lineage and philosophers who promote paramechanical views of the attitude of belief (including both realists and eliminativists who take beliefs—if they exist—to be theoretically posited cogs in cognitive machines). In this

chapter, I intend to make the deep disagreement explicit by arguing that, contrary to appearances, Ryleanisms and paramechanical views do not offer distinct accounts of a particular, agreed upon object of inquiry. Instead, Ryleans (ought to) countenance two distinct objects of inquiry where paramechanists countenance only one.[34]

The fundamental disagreement between Ryleans and paramechanists revolves around the distinction—upheld by the Rylean but collapsed by the paramechanist—between attitudes of belief and cognitive states of belief. On the paramechanical view, attitudes of belief are nothing more than theoretically posited cogs in cognitive systems. According to paramechanists, these particular human kinds—beliefs in the fidelity of husbands, the existence of gods, and the yellowness of applesauce—also happen to be paradigmatic natural kinds that play functional roles in cognitive systems. By the Rylean's lights, the paramechanist makes a category mistake by presumptively

---

[34] In Chapter 1, my primary goal was to reproduce the views of dispositionalists and interpretivists as accurately and charitably as possible. When arguing for my version of Ryleanism over my predecessors' in Chapters 6 and 7, I will again concern myself with faithfully representing their views. However, my aim in this chapter is a bit different. The primary goal of this chapter is to argue that Ryleans rightly distinguish between attitudes of belief and cognitive states of belief. But this claim is an interpretive fudge, or, to put the same point sanguinely, a rational reconstruction, of the stances of some of the Ryleans reviewed in Chapter 1. Most Ryleans do not make a big deal of the distinction between attitudes and cognitive states of belief; indeed, I suspect that much of the standard skepticism about Ryleanism stems from Ryleans' failure to explicitly endorse and explain this distinction, and thereby clarify the stakes of their debates. When referring to particular Ryleans by name, I will still attempt to represent them as accurately and charitably as possible. However, because my main interest in this chapter lies in casting Ryleanism in the best light (rather than accurately representing my Rylean ancestors), I will sometimes invoke Ryleanism (or Ryleans) in general as having tenets that are not actually upheld by one or more Ryleans. (Ironically, the dispositionalists are more likely to recognize the distinction than the interpretivists. This is ironic because it is the interpretivism inherent in Ryleanism that primarily motivates the refusal to conflate attitudes with cognitive states.)

assimilating attitudes of belief to cognitive states of belief.[35] I will extol the merits of this diagnosis in Chapters 3 and 4. In the present chapter, my goal is more modest: to make the distinction between attitudes of belief and cognitive states of belief precise, in part by differentiating it from a dozen other distinctions between varieties of belief.

My distinction between attitudes of belief and cognitive states of belief is not utterly novel. It has, however, only been invoked darkly. In this chapter, I will endeavor to bring the distinction fully into the light. To begin, I should emphasize that the distinction between attitudes and cognitive states is upheld by neither realist nor eliminativist paramechanists about belief. Both realist paramechanists and eliminativist paramechanists agree that attitudes of belief—if they exist —are cognitive states of belief. They just disagree about whether or not cognitive states of belief (and therefore attitudes of belief) exist. In Chapter 1, I articulated Ryleanism as a rival to realist paramechanical views about belief, including paramechanical mind-brain identity theories and paramechanical functionalisms. But I also mentioned that Ryleans are free to admit that the cognitive states of belief posited by realist paramechanists might well exist as the psychological underpinnings of attitudes of belief. There is no contradiction here, but the appearance of contradiction warrants explaining away.

Ryleanism is a rival to paramechanical views only insofar as both are theories of *attitudes* of belief. In most ordinary contexts, when I remark that the skater believes the

---

[35] By rejecting the conflation of attitudes with cognitive states, Ryleans should not be taken to imply that attitudes of belief are not natural kinds; see section 3.9., below.

ice to be thin, I am referring to the skater's attitude: (a portion of) the mental profile that I attribute to the skater as an inhabitant of my social environment. According to the paramechanist, to have a particular attitude is to have the right sort of cog in one's cognitive system; for example, to have an attitude of belief might be to bear psychofunctional relations to a propositionally structured mental representation that serve to power one's cognitive system to act as if that representation were true.[36] On this paramechanical functionalist way of thinking, the mental profile people verdically attribute to the skater is exhaustively characterized only by way of an explanation of how the skater's cognitive system is functioning.

For paramechanists, then, *attitudes* of belief can be unproblematically conflated with *cognitive states* of belief. Paramechanists usually treat this conflation as unproblematic because they assume that to make everyday attributions of attitudes is to (at least tacitly) undertake a particular sort of theoretical enterprise. In particular, paramechanical functionalists take everyday attributions of belief to consist in the activation of an implicit proto-scientific folk theory concerning the representational capacities and causal powers underlying human behavior. If belief attribution is best

[36] Of course, paramechanical philosophers and other cognitive scientists disagree vehemently about how to characterize cognition itself, in addition to the role cognitive states of belief play therein. As Gary Hatfield argues, the common ground between "classical, connectionist, dynamic, ecological, embodied, embedded, enactive, and extended" cognitivists is the view that "cognition is information processing that explains intelligent behavior" (2014: 361). In addition to these variables, different paramechanists take attitudes/cognitive states of belief (to be posited) to play producing (Fodor 1975) or structuring (Davidson 1963; Dretske 1988) causal roles in that information processing.

construed as the activation of such a folk psychological theory, then the attitudes of belief that folk psychologists attribute to each other are (folk) theoretical posits. In short, according to this standard paramechanical way of thinking, attitudes of belief are nothing more than theoretically posited cognitive states of belief. If such cognitive states of belief do not exist, then, as paramechanical eliminativists argue, attitudes of belief do not really exist either.

Ryleans treat folk psychology as an altogether less theoretical enterprise, unconcerned with the deep causes of human behavior. For the Rylean, the skater's mental profile can be exhaustively characterized in terms of the skater's patterns of behavior, thoughts, and feelings, without reference to the cognitive states that function to produce behaviors, thoughts, or feelings. Belief attribution is better construed as a pragmatic practice of capturing patterns in which people live their lives than as the application of a proto-scientific theory about the productive causes of behavior. Indeed, perhaps the central tenet of the Rylean tradition is that to conflate attitudes and cognitive states is to make a category mistake.

Baker, for example, writes that "[to] replace intentional explanations and descriptions in terms of beliefs by 'deeper' physical [or computational] explanations and descriptions in terms of brain states [or computational roles] is just to change the subject, as Donald Davidson rightly pointed out" (Baker 2003: 184). In a similar spirit, Schwitzgebel notes that "[t]he term 'attitude' once meant posture" and "recommend[s] that we retain a sense of those etymological origins." For Schwitzgebel, "to have an

attitude is, at root, to live a certain way" (Schwitzgebel 2011: 3). According to minimal Ryleanism, to possess a particular attitude of belief is to possess a pattern of dispositions that an interpretive scheme designates as constitutive of that belief. To believe is thus to live a life interpretable as the life of a believer. This dispositionalist and interpretivist understanding of the attitude of belief is silent on the question of whether psychologists will one day unveil cognitive states reasonably labeled 'beliefs'. Ryleanism, in its minimal form, does not entail the denial that cognitive states of belief of the sort posited by mind-brain identity theorists and functionalists exist. Nor does it entail the denial that each attitude of belief exists only in virtue of a particular cognitive state of belief. It entails only that theorists ought to refrain from conflating attitudes of belief with cognitive states of belief, in the absence of compelling evidence concerning how attitudes are realized in cognitive architectures.

According to Ryleans, then, metaphysicians of belief have to reckon with two distinct objects of inquiry. Attitudes of belief are the patterns of living as if the world were some way routinely attributed to believers (qua objects) by belief attributors in everyday life. Cognitive states of belief are the cogs in cognitive systems that produce believers' (qua subjects) actions, thoughts, and feelings. The deep disagreement between Ryleans on the one hand and both paramechanical realists and paramechanical eliminativists on the other hand is just this: paramechanists hold that attitudes of belief can be unproblematically conflated with (theoretical) cogs that (theoretically) enable believers to act; Ryleans deny that this conflation is unproblematic. Ryleans therefore

deny that paramechanical theories of cognitive states of belief are *ipso facto* theories of attitudes of belief. Moreover, Ryleans offer various dispositionalist interpretivisms and interpretivist dispositionalisms as versions of the only viable theory of attitudes of belief that refrains from conflating its objects of inquiry with cognitive states.

2. Thirteen distinctions between varieties of belief

As a way of further exploring the nature of this disagreement, I am now going to differentiate the distinction between attitudes of belief and cognitive states of belief from a dozen other ways in which philosophers have distinguished between varieties of belief, in (very roughly) ascending order of relevance to the debate between Ryleans and paramechanists. The first—Tamar Gendler's distinction between belief and alief—has next to nothing to do with my distinction between attitudes of belief and cognitive states of belief.[37] The distinctions will gradually get warmer. The twelfth—the distinction between beliefs as occupants of what Wilfrid Sellars termed the 'manifest image' and 'scientific image', respectively—captures the spirit of the distinction between attitudes and cognitive states. But Sellars's distinction still differs from the Rylean distinction in the nitty gritty details, and seeing exactly how they differ will serve to make the deep disagreement between Ryleans and paramechanists precise.

[37] I will address the first several distinctions superficially, gliding over tricky questions about how to understand each distinction in favor of simply stating one canonical form of the distinction and then briefly explaining why that distinction fails to divide attitudes from cognitive states.

FIGURE 1: THIRTEEN DISTINCTIONS BETWEEN VARIETIES OF BELIEF

1) implicit belief vs. explicit belief

2) binary belief vs. degrees of belief

3) System 1 belief vs. System 2 belief

4) dispositional belief vs. dispositions to believe

5) dispositional belief vs. occurrent belief

6) belief vs. assent

7) computational belief vs. algorithmic belief

8) personal belief vs. subpersonal belief

9) normative belief vs. natural belief

10) first-order belief vs. higher-order belief

11) folk psychological belief vs. psychological belief

12) manifest image belief vs. scientific image belief

13) attitude of belief vs. cognitive state of belief

2.1. First distinction: implicit belief vs. explicit belief

Implicit attitudes are a hot topic in 21st century philosophy of mind and epistemology. Philosophers have been impressed by experimental work—centering on the Implicit Association Test (IAT)[38]—suggesting that people's empirically revealed attitudes do not always match up with their explicitly endorsed attitudes (Greenwald et al. 1998). For example, the IAT reveals that many people explicitly denounce racism

[38] https://implicit.harvard.edu

while implicitly associating people of other races with negative stereotypes.[39]

Among philosophers, Tamar Gendler (2008) has done the most to articulate the distinction between explicit and implicit beliefs. Gendler proposes that these are two entirely distinct types of mental state. Belief is the type of explicit mental state revealed by the denunciation of racism; "alief"—"so-called because alief is *a*ssociative, *a*ction-generating, *a*ffect-laden, *a*rational, *a*utomatic, *a*gnostic with respect to its content, shared with *a*nimals, and developmentally and conceptually *a*ntecedent to other cognitive attitudes" (8)—is the type of implicit mental state revealed by the IAT. Gendler is a committed paramechanist, and dutifully conflates the attitudes attributed by belief attributors (and revealed by the IAT) with cognitive states of believers (and test-takers). For Gendler, then, the attitude of belief in question is whichever cognitive state produces explicitly anti-racist behaviors, while the alief in question is whichever cognitive state produces implicitly racist behaviors.

The fact that Gendler countenances both alief and belief in a paramechanist framework demonstrates that this distinction fails to serve my purposes. Whether or not Gendler is right that alief and belief are properly thought of as distinct cognitive states, attitudes of belief might well be fruitfully divided into explicitly endorsed beliefs and

---

[39] Implicit bias has recently undercome increased scrutiny, with studies suggesting that implicit bias (as measured by the IAT) correlates weakly with biased behavior (Oswald et al. 2013; Forscher et al. 2016). However this scientific debate—which concerns the instruments for measuring implicit attitudes as much as the existence of implicit attitudes—resolves itself will be of interest for both paramechanists and Ryleans.

implicit beliefs.[40] Even attitudes of belief that exist in relation to the believer herself may be implicit. For the Rylean, beliefs consist in patterns of dispositions that an interpretive scheme picks out as constitutive of belief. The implicit sexist may well wield an interpretive scheme according to which her own dispositions realize an implicit sexist belief without recognizing that she has the relevant dispositions, and thus without recognizing that she possesses (much less explicitly endorsing) the implicit belief in question.

2.2. Second distinction: binary belief vs. degrees of belief

Another recent trend revolves around the distinction between binary belief and degrees of belief. Bayesian philosophers find it useful to formalize the degree of belief a given individual has in a given proposition by means of a scale according to which '0' represents certainty that the proposition is false and '1' represents certainty that the proposition is true. According to most Bayesians, every individual has a degree of belief between 0 and 1 in every proposition they consider. It is unclear how this degree of belief relates to whether or not the individual flat out believes (or disbelieves) the proposition. The so-called 'Lockean Thesis' dictates that anyone with a degree of belief

[40] Or it might not. Schwitzgebel (2010) argues that we should embrace in-between belief (discussed in Chapter 1) rather than divide explicit belief from implicit belief. For Schwitzgebel, what we really believe about other races is a complex, messy result of combining our anti-racist behavioral dispositions with our racist behavioral dispositions. People who reveal racist associations via the IAT do not have two attitudes: one anti-racist belief and one racist alief. Instead, according to Schwitzgebel, they have a single (albeit complex and messy) attitude that lies somewhere in-between anti-racist belief and racist belief. See Chapter 6 for more.

over .5 flat out believes the relevant proposition, while anyone with a degree of belief under .5 flat out disbelieves the proposition. Some philosophers reject the Lockean Thesis in favor of the view that degrees of belief and binary beliefs, while both important states of believers, can come apart such that somebody can have a degree of belief of .99 (or .01) in a proposition but still disbelieve (or believe) that proposition. Yet other philosophers reject the idea that degrees of belief are quantitatively determinate, and treat ascriptions of degrees of belief as qualitative assessments of confidence rather than quantitative measurements. Finally, some philosophers think that only (quantitative or qualitative) degrees of belief exist, and that talk of binary belief should be replaced, wholesale, with talk of degrees of belief.[41]

This debate applies to both attitudes of belief and cognitive states of belief. Indeed, if the distinction between attitudes of belief and cognitive states of belief holds water, then the debate about whether belief is binary or degreed may well have two different resolutions. Cognitive states of belief might be binary—you either have the physical state or functional role that is the belief or not—or they might come in degrees—your connectionist architecture might comprise a range of degrees of belief depending on the strength of the relevant connections, for example. Alternatively, cognitive architectures might contain both binary states of belief and degrees of belief, as distinct cognitive states of belief. The same menu of options applies to attitudes of belief:

[41] Titelbaum (forthcoming: Chapter 1) provides a nice overview of this debate.

belief attributors might attribute just binary beliefs, just degrees of belief, or both.[42] In any case, it is not obvious that the correct answer to whether the cognitive state of belief is binary or degreed (or both) will be the same as the correct answer to whether the attitude of belief is binary or degreed (or both).

2.3. Third distinction: System 1 belief vs. System 2 belief

Two-systems (or dual-process) theories of cognition have also been extremely influential in philosophy over the last decade. One of their leading proponents, Jonathan Evans, notes that "all these theories have in common the distinction between cognitive processes that are fast, automatic, and unconscious and those that are slow, deliberative, and conscious" (Evans 2008: 255). A corollary distinction can be drawn between beliefs that arise from the fast, automatic, and unconscious System 1 and beliefs that arise from the slow, deliberative, and conscious System 2. For example, the belief that I see a knife

---

[42] This menu is not exhaustive. Indeed, I do not think any of these options adequately characterize attitudes of belief. To foreshadow, consider Schwitzgebel's notion of in-between believing. He notes, rightly, that neither binary belief nor degrees of belief (at least quantitatively construed) are nuanced enough to account for many quotidian belief attributions. Schwitzgebel writes that "Although some cases of in-between believing become manageable simply upon recognition of degrees of belief, cases such as those described above do not yield to this approach. It is not that Ellen and Geraldine simply have a low degree of confidence (say .6 on a scale from 0 to 1) in the truth of the proposition in question. Rather, they are disposed to feel in some situations quite confident in asserting one thing, while at the same time they are disposed to feel in other situations quite confident in asserting the opposite. Their doxastic condition is far from the kind of simple uncertainty that one might feel, for example, about the outcome of an election or the toss of a die. The cases that are the focus of this paper are no more manageable by an analog view of belief, on which belief is always smoothly describable by particular degrees of confidence, than by an all-or-nothing view of belief" (2002: 261). In Chapter 6, I will argue that while it is a step in the right direction, Schwitzgebel's notion of in-between believing still falls short of capturing the different styles in which people can hold the same belief.

before me likely arises from System 1: the perceptual mechanisms that produce this belief act swiftly and without my conscious control. The belief that the knife before me is a Bollock Dagger (due to the unique craftsmanship of its guard) requires some rational consideration, and thus likely arises from System 2.

The distinction between System 1 beliefs and System 2 beliefs may seem, prima facie, to be the same as the distinction between implicit beliefs and explicit beliefs. However, whereas the implicit/explicit distinction applies equally well to attitudes of belief and cognitive states of belief, both System 1 beliefs and System 2 beliefs are—by definition—kinds of cognitive state of belief. The distinction between System 1 and System 2 is a distinction between different kinds of cognitive processing. Attitudes of belief, insofar as they are distinct from cognitive states of belief, are not differentiated on the basis of the kinds of cognitive processing underlying the dispositions that make up the attitudes of belief.[43]

2.4. Fourth distinction: dispositional belief vs. dispositions to believe

Robert Audi (1994) has influentially distinguished between dispositional beliefs and dispositions to believe. Audi takes mathematical beliefs (and associated dispositions

[43] The distinction may not hold between cognitive states of belief either. As discussed below, it is plausible that cognitive states of belief exist only at a relatively abstract psychological level. In particular, cognitive states of belief might be individuated at a level of psychological explanation that abstracts away from mechanistic facts about whether the processing that produces cognitive states of belief is fast or slow. Thus, even if dual-process theories of cognition are viable, it is not obvious that this difference in processing would ground a corresponding difference in cognitive state of belief produced.

to believe) to serve as the paradigm illustration of this distinction. There are infinite beliefs that flow from my understanding of basic arithmetic. (I know that adding 1 to any number results in a greater number. So, for any n, I am disposed to believe that n+1 is greater than n.) But this does not mean that, right now, I possess the infinite beliefs that flow from my understanding of arithmetic. Instead, while I presumably possess the dispositional belief that 2 is greater than 1, I am merely disposed to believe that 273468 is greater than 273467, for example.

This distinction may be useful for both Ryleans and paramechanists. For the Rylean, I already have the behavioral, cognitive, and phenomenal dispositions that realize the pattern that is the (the attitude of) believing that 2 is greater than 1. However, on Audi's view I do not, before considering the mathematical relationship between 273468 and 273467, yet have the dispositions that amount to believing that the former is greater than the latter. I am, however, prone to picking up those dispositions when prompted by a relevant problem set: indeed, this propensity numbers among the dispositions that make up my belief that 2 is greater than 1.

For the paramechanical psychofunctionalist, meanwhile, I have the proposition that 2 is greater than 1 stored in my cognitive system—in a Fodorian Belief Box, say. I do not yet have the cognitive state of belief that 273468 is greater than 273467 stored in my cognitive system, but my cognitive system is situated such that I will quickly come to put the representation that 273468 is larger than 273467 in a Belief Box should the need

arise.[44] The distinction between dispositional belief and dispositions to believe can thus be wielded with regard to both attitudes of belief and cognitive states of belief.[45]

2.5. Fifth distinction: dispositional belief vs. occurrent belief

I have believed that basketball player Dikembe Mutombo's full name is Dikembe Mutombo Mpolondo Mukamba Jean-Jacques Wamutombo since I was a child, when I greatly enjoyed reciting "Dikembe Mutombo Mpolondo Mukamba Jean-Jacques Wamutombo" in singsong. Until last night, however, I had not had occasion to think about Mutombo's full name for many years.

Last night I was asked by a small child if "Dikembe Mutombo" is Dikembe Mutombo's real name. I gleefully responded that, as a matter of fact, his full name is "Dikembe Mutombo Mpolondo Mukamba Jean-Jacques Wamutombo." That I remembered Mutombo's full name shows that my *dispositional* belief has persisted through the years. I was always disposed to act in accordance with that belief, should

---

[44] Dispositions to believe are sometimes called 'implicit beliefs', as distinguished from beliefs (whether dispositional or occurrent a la the fifth distinction) that are (on the Rylean account) already dispositions of the believer or (on the paramechanist's account) explicitly represented. This is a distinct use of the term 'implicit belief' than is invoked by the first distinction; aliefs may well be explicitly represented in the relevant sense.

[45] With that said, Audi's distinction may be crucial to certain theories of cognitive states whereas it is optional for Rylean theories of attitudes. For Ryleans, it is perfectly plausible that I do already believe that 273468 is greater than 273467 before considering it, since I am already disposed to quickly assent to that proposition, infer that it is an instance of the general fact that n+1 is greater than 1, and feel superior to less rational animals when making that inference. On the other hand, the mind-brain identity theory and versions of functionalism that take cognitive states of belief to be discrete subpersonal states quickly run into the frame problem (Dennett 1978: 125) and have trouble for accounting how a cognitive system could possibly store so many dispositional beliefs (as opposed to mere dispositions to believe).

the occasion arise. Nevertheless, I lacked the *occurrent* belief that Mutombo's full name is Dikembe Mutombo Mpolondo Mukamba Jean-Jacques Wamutombo for the entire period between last night and the previous time I thought about Mutombo's full name (probably sometime in middle school).

Most paramechanists endorse this distinction between occurrent belief—the act of believing that *p* in the moment—and dispositional belief—the dormant, standing state of belief that *p.* Endorsing dispositional belief in this sense does not entail endorsing dispositionalism about belief. Both Ryleans and paramechanists hold that people have standing state beliefs that sometimes go a long time without being manifested in thought, feeling, or behavior. The debate between Ryleans and paramechanists revolves around the further question of whether attitudes of belief are patterns of dispositions to behave, think, and feel (whether or not those dispositions are manifested), as opposed to the cognitive states underlying those patterns (whether or not those cognitive states are activated).

For Ryleans, the distinction between dispositional belief and occurrent belief is something of a nonstarter anyway. Dispositionally believing *is* believing in the moment. There are, of course, various occurrent manifestations of dispositional beliefs. And the active thought—token of inner speech—that '*p*' is often a central manifestation of the belief that *p*. But Ryleans are quick to dismiss the idea that the occurrent token of inner speech 'the ice is thin' is itself a belief, as opposed to one of many possible manifestations of a belief. For instance, Ryle denied that the act of affirming the

proposition that 'the ice is thin' is any more central a manifestation of the belief that the ice is thin than the behavior of skirting the ice (1949: 134–135). For the Rylean, there is no such thing as nondispositional occurrent belief; there are just a range of occurrent manifestations of dispositional beliefs. I have occurrently believed that Mutombo's full name is Dikembe Mutombo Mpolondo Mukamba Jean-Jacques Wamutombo in every moment since I was little. I simply had not manifested that belief in behavior, thoughts, or feelings for a while. For the Rylean, then, the more apt distinction is between believing and manifesting one's belief.

2.6. Sixth distinction: belief vs. assent

Nevertheless, there is something to the idea that there are important differences between somebody who is merely disposed to act, think, and feel as if *p* were true and somebody who actively judges that *p* is true. Several philosophers amenable to a dispositionalist account of belief—and dissatisfied with the common distinction between dispositional belief and occurrent belief—have therefore instead made a distinction between dispositional belief and a complementary doxastic state that is actively (and usually consciously) held. Ronald de Sousa (1971) terms this latter state "assent", Dennett (1978, 1991) terms it "opinion", L. Jonathan Cohen (1992) terms it "acceptance", and Keith Frankish (2004) terms it "superbelief". Whereas belief is a pattern of dispositions (including, centrally, the dispositions to think—and, as Cohen rightly stresses, feel—that *p*), de Sousa writes that assent is an active "bet on truth for epistemic

ends" (1971: 61). Dennett ties de Sousa's notion more intimately to language, writing that "opinions are essentially bets on the truth of sentences in a language that you understand" (1998: 89). Cohen, in turn, ties Dennett's notion more intimately to practical reason, writing that "to accept that *p* is to have or adopt a policy of deeming, positing, or postulating that *p* — *i.e.* of including that proposition or rule among one's premisses for deciding what to do or think in a particular context, whether or not one feels it to be true that *p*" (1992: 4).

Frankish has provided the most in-depth treatment of the distinction between belief and assent to date. For de Sousa, Dennett, and Cohen, assent, opinion, and acceptance are to be sharply distinguished from belief. Frankish agrees that such a state of active assent to a sentence should be sharply distinguished from dispositional belief, but maintains that both are varieties of belief. Frankish argues for a two-strand theory of mind, each strand of which is centrally characterized by its own, proprietary form of belief. Dispositional belief is "non-conscious, possibly not subject to occurrent activation, partial [read: degreed], passively formed, probably non-verbal, and common to both humans and animals," while the other, assent-like, "superbelief" is "conscious, subject to occurrent activation, flat-out [read: binary], capable of being actively formed, often language-involving, and consequently, unique to humans and other language users" (2004: 23). Frankish thinks that both belief and superbelief are invoked in the everyday attribution of belief, and that we have good scientific reason to think humans are

accurately characterized as having both beliefs and superbeliefs.[46]

Frankish casts the distinction between belief and superbelief as a contribution to the paramechanical literature. He takes belief and superbelief both to be folk theoretical posits which, along with "two different types of mental processing … form two distinct levels of cognition" (2004: 4). Dennett, on the other hand, provides the distinction between belief and opinion as a tool for Ryleans. For Dennett, beliefs and opinions are distinct attitudes; both attitudes characterize language-wielding intentional systems, whereas only the former characterizes nonlinguistic intentional systems.[47] Hence our everyday practices of attributing articulate, linguistic opinions to our friends (but not our cats and dogs) while attributing inchoate, nonlinguistic beliefs to our cats and dogs (as well as our friends).

The upshot for our purposes is that the distinction at hand does not match up with the distinction between attitudes of belief and cognitive states of belief. Both Ryleans and paramechanists rightly appeal to the distinction between belief and assent.

2.7. Seventh distinction: computational belief vs. algorithmic belief

As noted, Frankish considers belief and superbelief to operate at "two distinct levels of cognition." Many philosophers and cognitive scientists have found it both

---

[46] The sixth and fifth distinctions most clearly differ insofar as assent (like dispositional belief) is merely *subject to* occurrent activation, rather than actually occurrently activated.
[47] However, see my discussion of the seventh and eighth distinctions, below, for some of the nuances of Dennett's position.

methodologically and metaphysically useful to distinguish between levels of cognition in this manner. Most influentially, David Marr (1982) distinguished between three levels of psychological explanation.[48] Cognitive systems can be described in terms of their functions (which Marr labeled 'the computational level'), in terms of the mechanisms ('the algorithmic level') that realize those functions, or in terms of the neural states or computer hardware (or whatever) that realize those mechanisms ('the implementational level'). Any given functional description of a cognitive system is neutral among multiple mechanical descriptions of that system, which are each in turn neutral among multiple neural descriptions of the system.

It may be tempting to identify attitudes of belief with specifications of belief at the computational level of psychological explanation. In particular, it might be tempting to understand patterns of dispositions to behave, think, and feel as person-level descriptions of how organisms function, abstracting away from the intricate subpersonal mechanisms realized by the nervous system.[49] But a Rylean description of an attitude of belief is not *ipso facto* a higher-level description of a cognitive state of belief, in the manner that a functional account of a cognitive state may *ipso facto* be a higher-level description of a mechanical state. Instead, the probing question raised by the Rylean tradition is whether a description of attitudes of belief is a description of the believer's cognitive system pitched at *any* level of psychological explanation, or whether it is a

---

[48] For influential versions of this distinction proposed by philosophers (though none are as influential as Marr's), see Dennett (1983), Millikan (1984), and Sterelny (1990).
[49] I will offer my full argument (barely adumbrated here) against this temptation in Chapter 4.3.3.

description of something else entirely. Do the folk normally attribute belief in order to help explain cognition, qua the production of intelligent behavior?[50] Do lay belief attributions directly posit cognitive functions or mechanisms at all?[51]

Ryleans have historically diverged on their answers to these questions. Dennett, takes the intentional stance to be a tool, in the first place, for the everyday prediction of behavior. But he also takes the intentional stance to be an indispensable tool for certain preliminary methods of cognitive psychological investigation (1987: 236–268).[52] Dennett thus sometimes equates intentional stance explanation—which usually invokes attitudes of belief—with the highest level of cognitive psychological explanation—in which case it instead invokes a purported cognitive state of belief at the most abstract level of psychological explanation. However, Dennett's commitment to belief-desire cognitive psychology as the highest level of cognitive psychological explanation is logically independent of his commitment to Ryleanism as a theory of attitudes of belief. Dennett could coherently give up the intentional stance as a tool in cognitive psychology while retaining his realism about attitudes of belief as picked out by the intentional stance as a tool for everyday social interaction.

Indeed, Ryle, Baker, and Mölder all endorse Ryleanism about attitudes of belief while vociferously denying that the lay attribution of belief has anything to do with the

---

[50] I will argue that they do not in Chapters 4 and 5.
[51] I will argue that they do not in Chapter 3.
[52] As we will see shortly, this double use of the intentional stance underlies Dennett's distinction between personal and subpersonal levels of explanation.

study of cognitive systems. Ryle noted that "part of the purpose of [*The Concept of Mind*] has been to argue against the false notion that … 'the mind' is what is properly describable only in the technical terms proprietary to psychological research" (1949: 299–300), as well as the accompanying notion that mental state terms must pick out parts of cognitive systems. For Ryle, some mentalistic terms, like 'belief', will not figure in any level of responsible cognitive theorizing; they serve an entirely non-scientific purpose. Belief attributions serve not to dissect cognition but to contextualize the ways in which people live as inhabitants of our social spheres.

So Ryleans disagree amongst themselves about whether or not cognitive states of belief ought to be posited at any level of psychological theorizing. But they ought to agree that paramechanists are wrong to tie the reality and nature of attitudes of belief to the reality and nature of cognitive states of belief (articulated at any level of cognitive psychological explanation).

2.8. Eighth distinction: personal belief vs. subpersonal belief

Closely related to the distinction between levels of psychological explanation is the distinction between personal explanations and subpersonal explanations. In explicating his version of Ryleanism, Dennett intends the distinction between personal belief and subpersonal belief to do much of the work I intend for the distinction between attitudes of belief and cognitive states of belief. We are getting warmer, but we are not there yet. Although attitudes of belief are inherently personal, cognitive states of belief

may or may not be personal too.

To wit, Dennett first distinguished between the personal and subpersonal levels of explanation in order to challenge the then-orthodox idea that all psychological explanation involves the attribution of mental states to whole persons.[53] Dennett argued that, in addition to "the explanatory level of people and their sensations and activities", psychologists ought to countenance "the sub-personal level of brains and events in the nervous system" (1969: 93). For Dennett, it was sometimes useful to attribute beliefs to whole persons, and sometimes useful to attribute beliefs to various cognitive subsystems (for which Dennett cheerfully reappropriated the epithet 'homunculi').[54] Once again, other Ryleans have disagreed with Dennett on this point. Ryle, Davidson, and Baker have all plumped for the traditional philosophical view that belief is only properly attributed to whole persons. Baker, for instance, takes it to be analytic—deducible from any acceptable definition of 'belief'—that beliefs "are global states of whole persons" (1995: 187).

Baker's case is strong with regard to attitudes of belief. The attitudes of belief that humans attribute to each other in everyday life are attributed *to each other*, as whole persons. These person-level attitudes include implicit (and otherwise subconscious) beliefs. People attribute implicit sexism to the sexist herself, as opposed to some sub-routine of her nervous system, even if she has no idea that she lives by discriminatory

---

[53] Zoe Drayson (2014) lays out the history of the personal/subpersonal distinction.
[54] Dennett's homuncular functionalism has been picked up and expanded upon by paramechanist (teleo)functionalists such as Lycan (1987) and Sterelny (1990).

beliefs. Nevertheless, Baker's arguments do not indict (much less convict) Dennett's hypothesis that some cognitive states of belief are subpersonal. Attitudes of belief are personal. Cognitive states of belief might be personal, subpersonal, or both.[55]

2.9. Ninth distinction: normative belief vs. natural belief

Philosophers of mind and epistemologists have drawn several interesting distinctions between beliefs qua natural states and beliefs qua normative states.[56]

Attitudes of belief have an important normative dimension. We ought to follow William James's (1896) admonition—"Believe truth! Shun error!"—in part because we are nigh-constantly judged to be intelligent or stupid, learned or ignorant, wise or naive, on the basis of what we believe. As James stressed, there is an epistemic ethics of attitudes of belief; John McDowell (1986) has argued, borrowing a phrase from Wilfrid Sellars (1956), that attitudes of belief reside in the normative Space of Reasons. There are also distinctively moral norms surrounding attitudes of belief. It is wrong to believe that women are inferior to men! It is right and just to believe in equal pay for equal work! Several philosophers have persuasively argued that belief attributors are less concerned with predicting and explaining behavior than with discerning whether believers fulfill

---

[55] Baker does implicitly point towards one possible difference between the attitude of belief and cognitive state of belief via her insistence that the attitude of belief is *intrinsically* personal. But Baker does not thereby resolve the question of whether cognitive states of belief ought to be attributed to cognitive systems as a whole (as opposed to various cognitive subsystems).

[56] It will follow from my analysis that there are really two key distinctions: normative belief vs. non-normative belief and natural belief vs. non-natural belief. Tensions between and reconciliations of the natural and the normative will be explored further in Chapter 7.

these normative roles (Falvey 1999; Morton 2003; McGeer 2007; Andrews 2015; Bohl 2015).

It is altogether less obvious that norms, epistemic or ethical, apply to cognitive states of belief. Cognitive states are, perhaps, what we need to work on altering in order to produce (epistemically and ethically) desirable attitudes of belief. But it is arguably not the underlying cognitive stuff that is praise- or blameworthy, because cognitive states reside in the Space of Causes, not the Space of Reasons. (Indeed, we often think that facts about underlying cognitive machinery can excuse people from moral praise or blame.) According to this line of thinking, only intersubjectively accessible person-level attitudes of belief invite normative evaluation.

I do think that there is an important (though non-constitutive) normative dimension to attitudes of belief, which sets them apart from cognitive states of belief.[57] Nevertheless, it is much too fast to simply label attitudes of belief 'the normative kind of belief' and cognitive states of belief 'the non-normative kind of belief'. For one thing, the idea that we cannot praise or blame somebody for their underlying cognitive states is a substantive claim in need of argument. Moreover, even if cognitive states of belief do not invite epistemic or moral normative evaluation, they may well retain a normative dimension. For example, teleofunctionalists hold that beliefs aim at accomplishing a biological purpose that is set for them by natural selection. If teleofunctionalists are right

[57] Portions of Chapters 3, 4, 6, and 7 will be devoted to articulating the normative roles that attitudes of belief play in human communities.

about the proper function of believing, then a certain kind of normativity is constitutive of cognitive states of belief.[58]

At the same time, it might be tempting to think of cognitive states of belief as natural—meaning, in this context, amenable to natural scientific inquiry—in a way that attitudes of belief are not. Cognitive states of belief are, after all, situated squarely in the Space of Causes: they are cogs in cognitive machines. Cognitive states of belief are also, relatedly, things that might be selected for in evolution by natural selection. Jerry Fodor (1989) considers his paramechanical project to involve "naturalizing the mind" because he thinks that attitudes of belief only qualify as natural properties if they are cognitive states of belief. The Churchlands are paramechanical eliminativists about attitudes of belief because they agree with Fodor on this point (as committed reductive naturalists), but doubt that cognitive states of belief will be vindicated by neuroscientific research. So, insofar as attitudes of belief are distinct from cognitive states of belief, one might think that attitudes are not amenable to naturalistic inquiry. After all, if attitudes of belief are not cogs in cognitive systems, then are they not also causally impotent? And if they are causally impotent, how could they be products of natural selection? Perhaps attitudes of belief are proper objects of inquiry for epistemologists, ethicists, and other humanists, but only cognitive states of belief are proper objects of inquiry for cognitive scientists. Both Ryle and Baker have adopted this view.

---

[58] Indeed, epistemic and even moral normativity *might* follow from the sort of normativity inherent in teleofunctions (Nolfi 2015), though I am skeptical that evolution grounds normativity in the relevant sense (see fn. 132).

Though I endorse a Rylean account of belief, I reject Ryle's desire for sharp provincial boundaries between philosophy and the sciences of the mind. Attitudes of belief are just as natural—just as amenable to natural scientific inquiry—as cognitive states of belief. In Chapter 4, I will make the case that attitudes of belief can be usefully thought of as indirectly causally potent, and naturally selected for, without being conflated with cognitive states of belief, even while residing in the Space of Reasons (in the same way that many philosophers think of perceivable colors as indirectly causally potent, and naturally selected for, without conflating them with physical properties of colored objects). Attitudes of belief, like perceivable colors, can be fruitfully studied as normatively characterized natural properties of people, considered in their role as objects in belief attributor's environments. A major source of evidence for this claim comes from the fecund research programs investigating the mindreading and metacognitive capacities of various animals.

2.10. Tenth distinction: first-order belief vs. higher-order belief

'Metacognition' is a notoriously fraught term across the cognitive sciences. Many psychologists use the term to refer to any thinking about thinking. Others use it to refer to the feelings generated by thinking, such as tip-of-the-tongue states (Proust 2015).[59]

---

[59] Proust (2015) provides an overview of the debate between 'inclusivists'—who take 'metacognition' to refer to any thinking about thinking—and 'exclusivists'—who take 'metacognition' to refer specifically to noetic feelings that enable organisms to evaluate and control their own thought processes.

Educational psychologists often use 'metacognition' to refer more specifically to "higher order thinking which involves active control over the cognitive processes engaged in learning" (Livingston 1997: 1). Comparative psychologists sometimes use 'metacognition' to refer to self-knowledge in general or, more specifically, knowledge about what one knows (Terrace & Son 2009). Othertimes, comparative psychologists use 'metacognition' in a sense more like the educational psychologist's: asking whether, for example, "nonhuman animals share humans' capacity for metacognition—that is, for monitoring or regulating their own cognitive states" (Smith 2009: 389). Our tenth distinction, between first-order belief and higher-order belief, can be read as invoking yet another notion of metacognition. In particular, it distinguishes first-order beliefs—beliefs about any part of the world other than beliefs—from higher-order, metacognitive beliefs—beliefs about other beliefs.

Attitudes of belief are characterized, in part, as being objects of higher-order thought: the interpretive schemes wielded by belief attributors. Nevertheless, the distinction between first-order and higher-order belief fails to capture the import of the distinction between attitudes of belief and cognitive states of belief. For starters, it is obviously not simply the case that attitudes of belief are first-order beliefs whereas cognitive states of belief are higher-order beliefs. Both attitudes of belief and cognitive states of belief can be either first- or higher-order: they can be about things that are not beliefs or they can be about other beliefs. Nor is it the case that attitudes of belief can be objects of higher-order belief whereas cognitive states of belief cannot. Cognitive states

of belief can be objects of higher-order beliefs too: I have beliefs about the cognitive state of belief that the beer is in the fridge, posited by Dretske (1991: 113). Attitudes of belief and cognitive states of belief are both subject to higher-order thought.

I am not going to use the term 'metacognition' in any of the ways canvassed at the start of this section. Instead, I am going to save the term 'metacognition' to stand for the first-person analogue of third-person mindreading. (I will use 'belief attribution' generically to refer to both the third-person and first-person cases). On this usage, metacognition is the quotidian act of construing oneself as believing something or other, just as mindreading is the quotidian act of construing somebody else as believing something or other. *This* notion of first- or third-person belief attribution cuts very close to the distinction between attitudes of belief and cognitive states of belief. Attitudes of belief are objects of everyday acts of mindreading and metacognition. Paramechanists hold that attitudes of belief are posited as theoretical entities by these everyday acts, and are thus identical to cognitive states of belief. Ryleans demur, and prescribe distinct philosophical and scientific treatments of, on the one hand, the attitudes of belief posited by everyday mindreading and metacognition and, on the other hand, the cognitive states of belief posited by some research in cognitive science.

2.11. Eleventh distinction: folk psychological belief vs. psychological belief

This most recent characterization of the distinction between attitudes of belief and cognitive states of belief suggests another familiar distinction. Have I not just

asserted that attitudes of belief are the objects of folk psychology, whereas cognitive states of belief are the objects of scientific psychology?

There is much to be said concerning the debates about folk psychology that have driven so much of the literature in philosophy of mind and psychology over the last 40 years. These debates will be central themes of Chapters 3, 6, 7, and 8. In this chapter, I will refrain from going into detail about the literature(s) on folk psychology. Instead, I will simply note that if I were to accept the distinction between folk psychological belief and scientific psychological belief, I might be misunderstood as admitting three substantive assumptions that I reject.

First, the term 'folk *psychology*' seems to presuppose that folk psychology is a theoretical (and perhaps proto- or pseudoscientific) enterprise. Second, the act of distinguishing scientific psychology from folk psychology seems to presuppose that scientific psychology does not employ any of the same methods as ordinary belief attribution. Third, the distinction might be read to imply that attitudes of belief—'folk psychological beliefs'—are not themselves objects of scientific inquiry. I will discuss these assumptions in turn.

As we have seen, Ryleans frequently challenge the first assumption; Baker, for instance, insists that "beliefs are not theoretical entities" (1995: 21). I too want to avoid the implication that belief attribution is an inherently theoretical (much less pseudoscientific) enterprise. I also reject the implicit assimilation of the beliefs actually attributed via mindreading to the beliefs that the folk *theorize* they are attributing via

mindreading. In order to resist this assimilation, Dennett's contrast between "folk theory" and "folk craft" is worth quoting at length.

> Folk psychology is an extraordinarily powerful source of prediction. It is not just prodigiously powerful but remarkably easy for human beings to use. We are virtuoso exploiters of not so much a theory as a craft. That is, we might better call it a folk craft rather than a folk theory. The *theory* of folk psychology is the ideology about the craft, and there is lots of room, as anthropologists will remind us, for false ideology …
>
> I want to distinguish between craft and ideology, between what we learn to do, and what our mothers and others have actually *told* us the craft was all about when they enunciated the lore, for what the anthropologists tell us is that craft and ideology are often quite distinct …
>
> Fodor, for instance, looks at the craft of folk psychology and tries to come up with a theory about why it works. His theory is that it works because it's a good natural history. It is actually an account of what is going on in the head, he thinks. All the things that seem to be salient in the practice of the craft [e.g., belief] actually have their isomorphs and homomorphs [in] the head. So he comes up with what he calls 'intentional realism' … But that is just one theory

> about why folk psychology works [and] there are some pretty good reasons for thinking it's a bad theory. (1998: 81–84)

Dennett's contrast between folk craft and folk theory drives a wedge between quotidian belief attribution and (scientific *or folk*) psychological theories that use the objects of quotidian belief attribution models for their theoretical objects of inquiry. The mere fact that cognitive scientists like Fodor and the Churchlands theorize that the folk craft of mindreading posits directly causally efficacious cognitive states of belief does not show that it does. Nevertheless, the contrast between folk craft and folk theory does not show that attitudes of belief are distinct from cognitive states of belief either. Attitudes of belief and cognitive states of belief might be unproblematically conflated even though attitudes of belief are not attributed by way of an inherently theoretical enterprise. If we use the folk craft of mindreading to attribute beliefs qua cogs in cognitive systems (even without any implicit theoretical supposition that such cogs do exist in actual cognitive systems), then attitudes of belief are cognitive states of belief (even if they are not the strictly regimented cognitive states of belief countenanced by Fodorian folk psychological theory).

Second, attitudes of belief and cognitive states of belief might be distinct even though cognitive psychologists draw heavily on features of attitudes of belief in order to articulate features of cognitive states of belief. Ryleans reject the following three theses: that lay belief attribution is an inherently theoretical enterprise, that the beliefs

attributed by the folk craft of mindreading are unproblematically conflated with the beliefs attributed by folk psychological theorists, and that attitudes of belief are unproblematically conflated with cognitive states of belief. These three theses do not require that Ryleans reject the further thesis that psychological theories can fruitfully rely on the resources provided by the folk craft of mindreading.[60] Heidi Maibom (2003) and Peter Godfrey-Smith (2005) have proposed model-theoretic accounts of folk psychology, according to which folk psychological models can be used in the service of either the folk craft of mindreading or scientific psychological theorizing. Model theories are compatible with Ryleanism, so long as lay people do not ordinarily construe the attitudes of belief they model as cognitive states of belief (when using them in the service of the folk craft of mindreading).[61]

Third, and most importantly, it bears repeating that attitudes of belief are as much objects of scientific inquiry as cognitive states of belief. Echoing a common refrain, Jaegwon Kim writes that "scientific properties must be cognitively invariant across different perceivers and cognizers" (2003: 95). If this is right, then it might be thought that Rylean attitudes of belief are not scientific properties: at least on some Ryleanisms (including Dennett's and my own), attitudes of belief can be cognitively variant across

---

[60] On the contrary, cognitive psychology plainly does (and will continue to) draw on everyday mindreading in various ways, whether or not it ends up countenancing cognitive states of belief. Consider the following analogous claim: utterances in English are distinct from sentences in the Language of Thought even though Fodorian cognitive psychology draws heavily on features of utterances in English in order to articulate features of the Language of Thought.

[61] I will discuss the model-theoretic view in more detail in Chapter 3, and it will play a key role in the arguments for my relativistic version of Ryleanism in Chapters 6, 7, and 8.

different belief attributors. But this thought relies on a misconception of attitudes of belief qua scientific phenomena. According to Ryleanisms, attitudes of belief exist relative to belief attributors, and thus should be studied as properties that emerge from the relationship between believers and belief attributors. Although the skater's belief, full-stop, is cognitively variant across belief attributors, relativized beliefs—such as the skater's-belief-per-a-particular-belief attributor's-model-of-the-belief-that-the-ice-is-thin—is cognitively invariant across cognizers. Whether or not a believer has a particular belief relative to a particular belief attributor is perfectly objective: it is in principle "intersubjectively ascertainable on the basis of shareable observational data" (Kim 2003: 95). Scientifically investigating Rylean attitudes of belief simply requires studying belief attributors as well as believers.

This is the main reason that I want to resist casting the distinction between attitudes of belief and cognitive states of belief in terms of the distinction between folk psychological belief and scientific psychological belief. Once we rid ourselves of the notion that the folk craft of belief attribution is a theoretical enterprise, there may be no problem with terming attitudes of belief 'folk psychological beliefs'. But there would remain a problem with contrasting 'folk' psychological beliefs with 'scientific' psychological beliefs, given that attitudes of belief are a variety of belief that are fruitfully studied by natural and social scientists of many stripes (as I will discuss in Chapters 3, 7.5, and 8; cf. Hochstein 2016).

2.12. Twelfth distinction: manifest image belief vs. scientific image belief

Although there are scientific theories about attitudes of belief, Ryleans should take attitudes to be pre-theoretical in one important sense: they are not, in the first place, posited via a theoretical enterprise. Cognitive states of belief are plausibly theoretical in precisely this to-be-explained sense: they are, in the first place, posits of theories in cognitive science. I have already mentioned Sellars's (1956) distinction between the Space of Reasons and the Space of Causes. Our penultimate distinction is also due to Sellars (1962): the distinction between belief located in the 'manifest image of man-in-the-world'—at a first pass, belief as it bears on everyday human experience—and belief located in the 'scientific image of man-in-the-world'—at a first pass, belief as a theoretical posit.

For Sellars, the manifest image is not unscientific; nor are beliefs unscientific because they are properties of the manifest image. He wrote that

> the manifest image is, in an appropriate sense, itself a scientific image. It is not only disciplined and critical; it also makes use of those aspects of scientific method which might be lumped together under the heading 'correlational induction'. There is, however, one type of scientific reasoning which it, by stipulation, does *not* include, namely that which involves the postulation of imperceptible entities, and principles pertaining to them, to explain the behaviour of perceptible things … Indeed, what I have referred to as the

> 'scientific' image of man-in-the-world and contrasted with the 'manifest' image, might better be called the 'postulational' or 'theoretical' image (1962: 7).

Beliefs occupy the manifest (rather than scientific) image insofar as they are manifest in experience (rather than theoretically posited). Some beliefs may, nevertheless, be made manifest in experience due to rigorous scientific investigation.

Sellars's distinction does not line up perfectly with my distinction between attitudes of belief and cognitive states of belief. Attitudes of belief exist in the manifest image, but so do cognitive states of belief, at least potentially. After all, if the most optimistic realists about cognitive states of belief are correct, then the advance of cognitive psychology and/or neuroscience will involve the transition of cognitive states of belief from theoretical posits to objects of experience. (We may be able, for example, to point at an fMRI brain scan of a wary skater and say "there! that is the belief that the ice is thin!") Similarly, although attitudes of belief are not (in the first place) posits of the scientific image, they do nevertheless fall within the explanatory and descriptive scope of the scientific image.[62] Sellars's distinction points us in the right direction insofar as it highlights the importance of the fact that attitudes of belief are manifest in experience. But it is not the distinction we are after. Instead, the distinction between attitudes of

[62] Cognitive scientists do, for example, posit a range of theoretical mindreading mechanisms that underlie the attribution (and thus existence) of attitudes of belief, as discussed in Chapter 3. Scientific theories of mindreading can be used to explain and describe attitudes of belief, as inhabitants of the manifest image.

belief and cognitive states of belief is a distinction between the variety of belief that exists *only* insofar as it is part of the manifest image and the variety of belief that stands or falls independently of the reality of the manifest image.

Paramechanists conceive the philosophical debate about the nature and existence of belief to be a debate about whether or not the beliefs that populate the manifest image are vindicated *in virtue of* also populating the scientific image. Paramechanical realists claim vindication because they foresee the beliefs posited in the manifest image being retained by the scientific image. Paramechanical eliminativists disagree. They agree with Sellars in holding that "man's conception of himself in the world does not easily accommodate the scientific image; that there is a genuine tension between them; that man is not the sort of thing he conceives himself to be; that his existence is in some measure built around error." Sellars found a predecessor of this view in

> Spinoza, who contrasted man as he falsely conceives himself to be with man as he discovers himself to be in the scientific enterprise. It might well be said that Spinoza drew a distinction between a 'manifest' and a 'scientific' image of man, rejecting the former as false and accepting the latter as true. (1962: 8)

It is easy to see the Churchlands as successors to Spinoza in this respect. However, in following Spinoza, they diverge sharply from their mentor Sellars. For Sellars noted that "if in Spinoza's account, the scientific image, as he interprets it, dominates the

stereoscopic view (the manifest image appearing as a tracery of explainable error), the very fact that I use the analogy of stereoscopic vision implies that as I see it the manifest image is not overwhelmed in the synthesis" (1962: 8–9). Sellars did not think the scientific image would replace the manifest image; he thought both images are veridical images of man-in-the-world, and that the task of philosophy is to show how the images can be fused without either image being reduced to the other.[63]

Unlike many of Sellars's other devotees (including not only eliminativists like the Churchlands but also paramechanical realists such as Millikan), I am devotedly Sellarsian on this point. Paramechanists make the mistake of conflating attitudes of belief with cognitive states of belief because they make the deeper mistake of attempting to either vindicate or eliminate the beliefs of the manifest image in virtue of facts about the scientific image. Sellars, on the other hand, recognized the need to describe and explain the manifest image by way of the scientific image without allowing that the entities, properties and activities populating the manifest image are cleanly reducible to the entities, properties and events populating the scientific image. The attitudes of belief that populate the manifest image of man-in-the-world are neither eliminable nor in need of vindication by way of the scientific image of man-in-the-world. Nevertheless, the

---

[63] Some Ryleans (certainly Baker and Mölder, and perhaps Davidson) are naturally read as advocates of what Sellars calls 'the perennial philosophy': roughly, the view that the scientific image must be assimilated to the manifest image. The perennial philosophy is something like the mirror image of the scientism of the paramechanist. But endorsing Ryleanism does not mean endorsing the perennial philosophy. Instead, the Rylean can (and I intend to) follow Sellars in insisting that the manifest image and scientific image can be stereoscopically fused (without one being reduced into the other).

scientific image can be taken advantage of to describe and explain what attitudes of belief are and why they exist.[64]

1.13. Thirteenth distinction: attitude of belief vs. cognitive state of belief

The deep disagreement between paramechanists and Ryleans comes down to the recognition, on the part of Ryleans, that attitudes of belief cannot be accounted for except as properties of the manifest image. Attitudes of belief exist only insofar as somebody recognizes that they exist.[65] Cognitive states of belief exist (or fail to exist) whether or not anybody recognizes that they exist. Ryleans hold, on that basis, that attitudes of belief must be distinguished from cognitive states of belief.

Debates about whether or not cognitive states of belief exist pertain to the question of how agents are wired such that they are able to think and act in the world.

---

[64] Sellars did admit that the scientific image is, in one sense, explanatorily prior to the manifest image, but he did not mean that the manifest image is reducible to the scientific image. Instead, he meant that "the scientific image [can] recreate in its own terms the sensations, images, and feelings of the manifest image" (38). In other words, he meant that the scientific image can be used to explain and describe the manifest image (whereas the manifest image can—at best—only incompletely explain and describe the scientific image). But Sellars thought that there is more to philosophy—more to "understand[ing] how things in the broadest possible sense of the term hang together in the broadest possible sense of the term" (1)—than explaining and describing. There is also a dimension of normative understanding that can only be achieved via active participation in a community governed by norms. Although the scientific image is primary in the dimension of describing and explaining, the manifest image is all we have in the dimension of knowing how to engage in normative practices. On those grounds, Sellars concluded that the manifest image is irreducible to the scientific image, and that the one true image of man-in-the-world is the projected synoptic image afforded by fusing the manifest and scientific images.

[65] In particular, as I will discuss in Chapter 5.2., attitudes of belief exist only insofar as somebody "outer recognizes" (Haugeland 1998: 285) that they exist—that is, would recognize the relevant patterns of dispositions as beliefs. The occurrent recognition that a particular believer actually has the relevant pattern of dispositions is not required to render that pattern a belief.

Ryleans regard this question as irrelevant—or only contingently relevant—to the question of whether or not certain organisms have attitudes of belief, because whether or not an organism has an attitude of belief is a question about the role that organism plays—or would play—in the experience of belief attributors. Ryleans therefore hold that the beliefs people attribute to each other in everyday life—qua attitudes of organisms (considered as objects)—should not be conflated with cognitive states of organisms (considered as subjects). Moreover, they hold that findings about the cognitive states of organisms are, at most, only contingently relevant to the ontological status of those organisms' attitudes.

As I will discuss in Chapter 4, color scientists have found that objects have no physical properties that are completely isomorphic with perceivable color. But this finding does nothing to impugn the ontological status of color qua property of objects of experience. So too, Ryleans hold that even if cognitive scientists find that there is nothing in brains (or cognitive systems functionally described) that is isomorphic to any given attitude of belief, this finding would do nothing to impugn the ontological status of that belief qua mindreadable and metacognizable property of a believer.

3. Ryleans and representationalists as bedfellows

In 21st century discussions, Ryleanism about belief is usually cast as incompatible with representationalism about belief. The distinction between attitudes of belief and cognitive states of belief dissolves this apparent conflict. If we resist the paramechanical

urge to conflate attitudes with cognitive states, then it becomes clear that Ryleanism is an account of the former and representationalism (combined with psychofunctionalism or another theory of subjects' relations to mental representations) is an account of the latter.[66] As accounts of distinct phenomena, they are perfectly compatible.

Consider Jake Quilty-Dunn and Eric Mandelbaum's recent brief against Schwitzgebel's brand of Ryleanism. Quilty-Dunn and Mandelbaum present powerful arguments for various instantiations of the conclusion that their preferred theory of belief—a psychofunctional representationalism featuring fragmented belief storage—is explanatory in ways that Ryleanism is not. Psychofunctional representationalism explains how beliefs are "elements of causal-mechanical explanations of behavior" (2017: fn. 9); Ryleanism "does not posit concrete mental particulars that causally interact according to psychological laws" (§3.1). Psychofunctional representationalism explains—by positing mechanisms behind—sorting behaviors, Frege cases, the truth evaluability of beliefs, the distinction between representational content and representational vehicle, dissonance, belief change, and inference (§§3.2–3.6). For example, dissonance—the phenomenon of disconfirming information causing negative valence—comes about "because beliefs are concrete particulars with causal powers, governed by, inter alia, the laws of dissonance" (§3.5). (This may sound unsatisfying,

[66] My denial that attitudes of belief are psychofunctional relations to mental representations—representations understood as cogs in cognitive/representational systems—should not be read to imply the denial that attitudes of belief come equipped with intentional content—and are thus 'representational' in another standard sense. I discuss how Rylean attitudes of belief represent the world in Chapter 7, Sections 7.2.2., 7.2.6, and 7.2.7.

except that psychological experimentation can unveil these laws.) Ryleanism does not offer rival explanatory mechanisms that might enact these laws.

I agree with Quilty-Dunn and Mandelbaum that psychofunctional representationalism fulfills these explanatory ends, whereas Ryleanism does not. Anti-representationalist theories of cognitive architecture (Chemero 2009) better fulfill these ends than Ryleanism too. So do representationalisms that favor teleofunctional accounts of the relation between subjects and mental representations over psychofunctional accounts (Neander 2017). This explanatory lack is not due to Ryleanism being a bad theory of cognitive states of belief; it is due to Ryleanism being an account of something else entirely.[67] Quilty-Dunn and Mandelbaum write that "a desideratum for a theory of belief is to explain how beliefs [productively] cause behavior" (2017: §3.1). And it is clearly a desideratum for theories of cognitive states of belief. However, it need not be a desideratum for theories of attitudes of belief, which (I will argue in Chapter 3) should not be assumed to be inner causes that produce behavior.[68]

If attitudes are distinct from cognitive states, then representationalism (or any other theory of cognitive architecture) and Ryleanism may well both be true: psychofunctional relations to mental representations may be the cognitive categorical

---

[67] This point should be taken as a criticism of Schwitzgebel—who fails to distinguish between attitudes of belief and cognitive states of belief, and thereby casts his Ryleanism as a rival to representationalism—as much as a criticism of Quilty-Dunn and Mandelbaum.

[68] As I will argue in Chapters 4 and 5, accounts of attitudes of belief should explain how they function as aspects of belief attributors' social environments, rather than believers' cognitive systems. (Chapters 6 and 7 will then undertake this explanatory work in earnest.)

bases of the patterns of dispositions that lay attributors identify with beliefs. I will spend Chapters 3 and 4 arguing that we have good reason to refuse to conflate attitudes with cognitive states. If my arguments are sound, one happy consequence will be some perspective on how our best theories of belief might all be laying their hands on different parts of the same elephant.[69]

4. Fifteen theoretical alternatives concerning the metaphysics of belief

Ryleans need not be disinterested in cognitive states of belief. While discussing the relationship between attitudes and cognitive states, Peter Godfrey-Smith wisely echoes Sellars's conception of philosophy in stressing that "one of the roles for philosophy … is to describe the coordination between the facts about interpretations and the facts about wirings-and-connections" (2004: 149). In other words, philosophers should not only give a metaphysics of attitudes of belief and a distinct metaphysics of cognitive states of belief. They should also explain the connections between the two. I wholeheartedly agree with Godfrey-Smith, though (some scattered remarks notwithstanding) I will not take up the project he recommends in these pages. Instead, I

[69] Unlike Fodorian psychofunctional representationalism, Quilty-Dunn and Mandelbaumian psychofunctional representationalism does not rely on the assumption that attitudes are unproblematically conflated with cognitive states. (They do conflate the two, but this conflation is a vestigial assumption rather than working premise in their arguments.) Indeed, insofar as they stress that the "generalizations about belief [established by cognitive science] are counterintuitive" (2017: §3.5), Quilty-Dunn and Mandelbaumian representationalism might draw strength from the distinction (whereas Fodor thinks psychological laws mostly affirm what his Granny already knew).

will focus on developing and defending a Rylean metaphysics of attitudes of belief. I hope that this metaphysics of attitudes of belief will be illuminating in its own right. But I also hope to enable more clearheaded accounts of the nature of cognitive states of belief, as well as more nuanced and realistic accounts of how attitudes of belief and cognitive states of belief are coordinated.

The Rylean tradition highlights the possibility that attitudes of belief and cognitive states of belief deserve separate analyses. This possibility should lead the systematic philosopher of belief to consider the following fifteen generic options.

FIGURE 2: FIFTEEN THEORETICAL ALTERNATIVES ON THE METAPHYSICS OF BELIEF

Five Paramechanical Options

1. Eliminativism about attitudes/cognitive states (Rosenberg 2015).

2. Mind-brain identity theory about attitudes/cognitive states (Polger & Shapiro 2016).

3. Computational functionalism about attitudes/cognitive states (Fodor 1987).

4. Pure functionalism about attitudes/cognitive states (Jackson & Pettit 1990).

5. Teleofunctionalism about attitudes/cognitive states (Dretske 2000).

Six Rylean Options

6. Ryleanism about attitudes; eliminativism about cognitive states (Baker 1995).

7. Ryleanism about attitudes; mind-brain identity theory about cognitive states (Place 1958).

8. Ryleanism about attitudes; computational functionalism about cognitive states (Dennett 1989).

9. Ryleanism about attitudes; pure functionalism about cognitive states (Davidson 1963).

10. Ryleanism about attitudes; teleofunctionalism about cognitive states (Dennett 2017).

11. Ryleanism about attitudes; agnosticism about cognitive states (Ryle 1949).

Four Attitude-Skeptic Options

12. Eliminativism about attitudes; mind-brain identity theory about cognitive states (Crick 1995).

13. Eliminativism about attitudes; computational functionalism about cognitive states (Piccinini 2006).

14. Eliminativism about attitudes; teleofunctionalism about cognitive states (Lycan 1987).

15. Eliminativism about attitudes; agnosticism about cognitive states (Stich 1997).

When I argue against the mind-brain identity theory and various functionalisms in Chapter 4, I will argue against them as paramechanical theories of attitudes of belief.

Traditionally, Ryleans have differed on how they treat cognitive states of belief—when they recognize the distinction between attitudes and cognitive states at all. Some Ryleans have doubted that any cognitive states resemble attitudes of belief sufficiently to warrant the title 'belief'. Dennett, for example, writes that

> I am as staunch a realist as anyone about those core information-storing elements in the brain, whatever they turn out to be, to which our intentional interpretations are anchored. I just doubt that those elements, once individuated, will be recognizable as the beliefs we purport to distinguish in folk psychology.

(Dennett 1987: 71)

Dennett thus oscillates between eliminativism, computational functionalism, and teleofunctionalism about cognitive states of belief. Baker does not think Dennett's rhetoric is stark enough. She writes that

> [it] is as if Dennett were saying: 'The intentional stance is fine if you don't need the more accurate understanding that you can get from the design or physical stances.' That's like saying that looking at the Mona Lisa from 5 ft away is fine if you don't need the more accurate understanding that you can get from a microscope; but, of course, you see it better with a microscope. This would be a joke. You can see the paint better, but you do not see the Mona Lisa better through a microscope. You lose the Mona Lisa altogether. What you see by ordinary observation of the Mona Lisa is not just a summary of brush strokes. You can't just drop down a level for a better view. The patterns seen from the intentional stance are not just a kind of rough sketch of what is really going on [at] the level of the physical stance. (Baker 2003: 190)

For Baker's money, cognitive scientists will find nothing remotely like beliefs—nothing that does anything resembling the work beliefs do—when looking at fMRI scans, just as we find nothing remotely like the Mona Lisa—nothing that does anything resembling

the work the Mona Lisa does—when looking through a microscope. Other Ryleans are more ecumenical. Schwitzgebel allows that "[s]omeday, an advanced neuroscience may discover, corresponding to each of our beliefs, specific neural bases of the sort envisioned in the most optimistic type-type reductionism" (Schwitzgebel 2002: 270). In the meantime, though, Schwitzgebel stresses that scientists have not confirmed the existence of cognitive states of belief at the computational or algorithmic levels of psychological explanation, much less successfully reduced beliefs to states of the nervous system.

I remain agnostic about whether or not there are cognitive states that correlate with attitudes sufficiently to deserve to be titled 'beliefs'.[70] We will just have to wait and see whether BELIEF ends up being an indispensable concept for cognitive psychology and cognitive neuroscience as those sciences mature. If a mature and persuasive science of cognitive architecture posits cognitive states deserving of the title 'belief'—perhaps as psychofunctional relations to mental representations stored in fragments, as per Quilty-Dunn and Mandelbaum—then I will adopt the posited metaphysics of cognitive states of belief accordingly. If the eliminativist wager turns out to be the smart money, and the best science dispenses with reference to beliefs qua cogs in cognitive machines, then I will be happy to do without cognitive states of belief in my ontology. Either way, I will remain a dispositionalist and interpretivist about attitudes of belief. The ontological

[70] See Stich (1997) for a fairminded (if not disinterested) essay on this topic; my own views are further spelled out in the replies to objections in Chapter 5.

status of attitudes of belief is not fatefully tied to the eventual elimination or vindication of cognitive states of belief. No matter how our minds and brains work, we are believers because we possess patterns of dispositions that interpretive schemes designate as constitutive of belief. In support of this Rylean thesis, I will spend the next two chapters arguing directly against the standard paramechanical conflation of attitudes of belief with cognitive states of belief.[71]

5. The argument to come

In summary form, my argument that attitudes of belief ought not be conflated with cognitive states of belief will proceed as follows.

P1. In providing an account of a phenomenon, we ought not conflate it with another phenomenon if we have good reason to doubt that the phenomena are type-identical and can provide a satisfactory account of the former phenomenon without conflating it with the latter phenomenon.

P2. We have good reason to doubt that attitudes of belief and cognitive states of belief are type-identical.

P3. We can provide a satisfactory account of attitudes of belief without conflating

[71] Terminological note: from here on out, I will sometimes use the unadorned term 'beliefs' to refer to attitudes of belief. Similarly, I will sometimes use the term 'believers' to refer to creatures possessing attitudes of belief. I will never use 'beliefs' without qualification to refer to cognitive states of belief.

them with cognitive states of belief.

C. In providing an account of attitudes of belief, we ought not conflate attitudes of belief with cognitive states of belief.

Before rising to the defense of these premises, I would like to point out how Ryleanism and a commitment to the distinction between attitudes and cognitive states are mutually reinforcing. The existence of Ryleanism—a satisfactory account of attitudes of belief that does not conflate them with cognitive states—renders P3 of the argument for the distinction true. In turn, assimilating the conclusion of that argument should drive philosophers to adopt (at least minimal) Ryleanism, since—as argued at the end of Chapter 1—Ryleanism is the only going account of attitudes of belief that does not conflate them with cognitive states of belief.

P2 of the argument is where the real action is; the primary mission of the remainder of Part I will be to drum up doubt that attitudes of belief and cognitive states of belief are type-identical. Most philosophers of mind and psychologists presume the type-identity of attitudes of belief and cognitive states of belief, because they presume that lay belief attributors ordinarily construe beliefs as inner causes that produce behavior—and thus as cognitive states of belief. In Chapter 3, I will deny that there are any objective grounds for this dogma, while mustering emerging empirical evidence in support of the alternative, Rylean thesis that the folk ordinarily construe beliefs as patterns of living that contextualize behavior.

Chapter 3: Beliefs as inner causes?

1. Introduction

In 1949, Gilbert Ryle lambasted the "paramechanical" thesis that the everyday attribution of beliefs amounts to theoretically positing inner—"occult"—causes of behavior. In its place, Ryle proposed that the folk construe beliefs as patterns of living. When we say that *s* believes that *p*, we mean that *s* is disposed to act, react, think, and feel as if *p* "in a thousand different situations" (Ryle 1949: 44). By interpreting Kyrie Irving as believing that the Earth is flat, for example, I can fit what might be otherwise puzzling individual behaviors—such as eye-rolling during a total solar eclipse—into a coherent pattern of actions, thoughts, and feelings. His behavior makes sense given that he is prone to feeling distrustful of science teachers, noticing that the ocean appears endlessly level when he flies across the so-called 'globe', and so on. By attributing a belief to Kyrie, I can better understand one of my favorite basketball players without making any assumptions about his underlying psychological processes.

Ryle's dispositionalist account of lay belief attribution was popular—orthodox, even—for about a decade; it thrived alongside the influential philosophical psychologies of the later Ludwig Wittgenstein and his student Elizabeth Anscombe (1957). Since Donald Davidson published "Actions, Reasons, and Causes" in 1963, however, the consensus in philosophy of mind and psychology has turned firmly anti-Rylean.

According to Davidson, only causes that mechanically produce behavior serve to explain behavior. And, in the course of everyday life, people frequently attribute beliefs when explaining intentional behaviors. So, Davidson concluded that the folk construe beliefs as the inner causes—the cogs in cognitive systems—that produce the intentional behaviors they explain. I will call this thesis 'Davidson's dogma'.

Although the debate in action theory still rages (D'Oro & Sandis 2013), most contemporary philosophers of mind assume Davidson's dogma. It forms a key premise in both Jerry Fodor's (1987) argument for the existence of belief[72] and Paul Churchland's (1984) argument for the non-existence of belief.[73] Neither author offers new arguments in support of this premise. Churchland (1970: 214–215) thanks Davidson for establishing the dogma, and works out some of its consequences. Fodor simply "stipulates" that the folk construe beliefs as inner causes, explaining that "this seems to me intuitively plausible; if it doesn't seem intuitively plausible to you, so be it. Squabbling about intuitions strikes me as vulgar" (1987: 10).[74]

---

[72] Fodor: "commonsense psychology takes for granted that overt behavior comes at the end of a causal chain whose links are mental events—hence unobservable … we are all—quite literally, I expect—born mentalists and Realists; and we stay that way until common sense is driven out by bad philosophy" (1987: 7).

[73] Churchland: commonsense psychology is "a theory that postulates a range of internal states whose diverse causal relations are described by the theory's laws" (1984: 98).

[74] The practice of intuiting Davidson's dogma has not rescinded among philosophers of mind since the 1980s; cf. Nichols & Stich (2003: Chapter 1), Goldman (2006: Chapter 2), Egan (2014: 138), Mandelbaum (2014: footnote 68), Carruthers (2015a: Chapter 2); Crane (2016: 28–49); Quilty-Dunn & Mandelbaum (2017: Section 3.1).

I do not intend to squabble about intuitions. Instead, I will assess candidate objective grounds for Davidson's dogma. In Sections 3 and 4, I will examine two fecund scientific literatures on lay belief attribution. Psychologists studying theory of mind and behavior explanation often cite Davidson in support of their assumption that the folk construe beliefs as inner causes, but they provide no empirical evidence to support Davidson's dogma. Indeed, recent experimental findings better fit the Rylean view that the folk construe beliefs as patterns of living. Moreover, as I will argue in Section 5, Davidson's influential argument for his dogma is unsound. Against Davidson, beliefs can serve as reasons that uniquely explain actions without being construed as inner causes that produce those actions. Ryle's dispositionalism could already account for the supposedly mysterious connection that Davidson identified between beliefs and the actions they uniquely explain.

Cognitive scientists can (and do) use the beliefs that humans attribute to each other in everyday life as models for the cognitive causes of behavior (Godfrey-Smith 2005). Fodor might be right that such modeling is the only fruitful way to study cognition, or Churchland might be right that neuroscientists will prove Fodor myopic. Regardless, I intend to reveal that there is no extant justification, beyond philosophers' and psychologists' bare intuitions, for the assumption that the folk regularly construe beliefs as inner causes that produce behavior.

2. Unpacking Davidson's dogma

Before assessing the empirical evidence, it is worth getting a firmer grasp on what it would mean for the folk to construe beliefs as inner causes that produce behavior.

2.1. …beliefs as inner causes that produce behavior

In this statement of Davidson's dogma, the adjective 'inner' serves primarily to clarify the sense of the noun 'cause'. For a cause to be construed as inner, in the sense Ryle (1949: 19) criticized, it must be involved in an unobservable mental process that produces observable behavior. Strictly speaking, however, the unobservability of belief is not entailed by Davidson's dogma.[75] Instead, what all proponents of Davidson's dogma accept is that the folk posit beliefs as cogs in the mental (as opposed to physical) systems that generate behavior. Thus, in what follows, readers should take the term 'inner cause' to signify 'cog in a cognitive mechanism', with no prejudice as to observability.

In Davidson's own terminology, inner causes are the mental variety of 'producing causes' (1967: 703). Producing causation is itself but one variety of causation countenanced by folk reasoning. Ned Hall has persuasively argued that "our thinking

[75] For a recent argument against the observability of belief, see Spaulding (2015); for defenses, see Gallagher & Hutto (2008) and (at least for 'belief-like states') Herschbach (2015).

about causation recognizes two basic and fundamentally different varieties of causal relation": dependence and production (2004: 276). Dependence causation

> is simply that: counterfactual dependence between wholly distinct events. In this sense, event c is a cause of (distinct) event e just in case e depends on c; that is, just in case, had c not occurred, e would not have occurred. The second variety is rather more difficult to characterize, but we evoke it when we say of an event c that it helps to generate or bring about or produce another event e, and for that reason I call it 'production'. (Hall 2004: 225)

Ryle readily allowed that the folk construe intentional behaviors as being counterfactually dependent on beliefs. Just in case Kyrie did not believe that the Earth were flat, he would not have been thoroughly unimpressed by the total solar eclipse.[76]

---

[76] Philosophers offer various proposals about how to make this dependence relation precise. Famously, David Lewis (1973) argued that dependence should be understood via bare counterfactual about possible worlds: Kyrie's belief caused his eye-rolling just in case there is a possible world in which both Kyrie's belief and his eye-rolling obtain that is closer to the actual world than any possible world in which his belief obtains but his eye-rolling does not. James Woodward (2003) has since argued that our thinking about causation reflects our practical interests, proposing an interventionist account of dependence causation: Kyrie's belief caused his eye-rolling just in case manipulating his belief would have altered (or prevented) his eye-rolling. Woodward acknowledges that his interventionist account is "a counterfactual theory of causation, not in the [Lewisian] sense that [it] claim[s] to offer a reductive analysis of causal claims in terms of some (noncausal) notion of counterfactual dependence, but in the sense that [it] claim[s] that there is a systematic connection between causal claims and certain counterfactuals" (2003: 70). Moreover, Reutlinger (2012) argues that, because Woodward's theory concerns only metaphysically possible interventions rather than physically possible interventions (Woodward 2003: 10–11), it collapses into a bare counterfactual account of dependence causation á la Lewis.

The thesis that beliefs are dependence causes of behavior is uncontroversial.[77] Davidson's dogma entails the different, stronger claim that the folk construe beliefs as producing causes. In particular, according to Davidson's dogma the folk attribute beliefs as cognitive states that function to generate behavior via cognitive mechanisms.[78]

2.2. The folk construe…

Across cultures, people engage daily in the social practice of interpreting themselves and other people as believers. The psychologist Ian Apperly calls the folk ability to attribute beliefs (and other mental attitudes) "mindreading", instead of the more traditional 'theory of mind', because the latter

> implies that mindreading consists in having a theory about how the mind works.

---

[77] Two caveats. First, eliminativists would argue that beliefs—being non-existent—are not dependence causes of behavior. Second, some theorists might consider it a mistake to think of the explanatory relationship between beliefs and behavior as causal at all (Reutlinger 2017). I am agnostic about whether belief explanations always invoke causes of some sort; I am committed only to the conditional that if belief explanations are causal explanations, then they paradigmatically invoke dependence, not producing, causes. Because Ryle (1949: 122) and other Ryleans (Baker 1995; Dennett 1998) countenance belief explanations as invoking dependence causes, I adopt that working hypothesis.

[78] Whereas Lewis (1973) and Woodward (2003) have offered clear (if controversial) analyses of dependence, less has been done to unpack production. Davidson's (1967) view was that producing causation should be explicated in terms of laws of nature. Wesley Salmon (1984) and Phil Dowe (2000) argue that producing causation is the transfer of a conserved quantity (such as mass) from cause to effect, while Stuart Glennan (1996) maintains that claims about producing causation posit mechanisms. Analytic disagreements aside, these theorists agree that explanations of production unveil (in one way or another) the mechanism by which cause leads to effect. Accordingly, Kyrie's belief caused his confession just in case it was a cog in the cognitive mechanism that generated the rolling of his eyes.

> In fact this is, at best, just one theoretical possibility among many. And for my purposes it unhelpfully suggests that 'theory of mind' is something that one *has* rather than something that one *does*. (Apperly 2011: 3)

Apperly's contrast between actively engaging in mindreading and passively having a theory of mind is reminiscent of Daniel Dennett's earlier contrast (discussed in Chapter 2) between 'folk craft' and 'folk theory': "between what we learn to do, and what our mothers and others have actually told us the craft was all about when they enunciated the lore, for what the anthropologists tell us is that craft and ideology are often quite distinct" (Dennett 1998: 81). As Dennett insinuates, there may be a gulf between how ordinary folk craft and rarefied folk theory respectively construe belief.

While admitting that "there is little doubt that many folk subscribe to the idea that mental phenomena [including belief] are causal in a productive, mechanical way", Daniel Hutto speculates that this is "not because the folk have closely attended to what they do when using their folk concepts in practice" (2011: 141). Hutto suggests that folk theorists often construe belief as causally productive of behavior because they hold theoretical commitments to a 'mechanical/causal' worldview. In any case, the fact that Fodor's Granny (Fodor 1987: 6) has a pet theory that explicitly countenances Davidson's dogma does not entail that her folk craft—her cognitive ability to attribute belief—implicitly construes beliefs as causally productive of behavior. By arguing that the

everyday attribution of beliefs does not amount to positing occult causes, Ryle certainly did not intend to deny that ordinary folks had fallen for the theoretical allure of the myth of the ghost in the machine. He intended only to deny that the myth had infected people's folk craft of belief attribution.

Hutto might be wrong. Some philosophers claim that we have transparent conscious access to the concept of belief that we employ when mindreading. If they are right, then perhaps folk theory accurately recounts folk craft. Alternately, it is possible that the mechanical/causal folk theory exerts strong top-down influence on the folk craft, such that the mindreader ordinarily construes beliefs as inner causes precisely because of her explicit theoretical commitments about the nature of belief. I do not propose to attempt to settle the (largely empirical) question of the relationship between folk theory and folk craft in this chapter. Whatever the relationship between folk theory and folk craft, it is how belief is construed in crafty mindreading that is at stake in the debate over Davidson's dogma. After all, it is the folk craft of mindreading—the nearly universal human ability to attribute beliefs, desires and other mental attitudes—that subserves everyday belief explanations.

2.3. Davidson vs. Davidson's dogma

I have defined Davidson's dogma as portrayed in mainstream philosophy and psychology: as the thesis—widely credited to Davidson—that the folk craft of

mindreading construes beliefs, themselves, as producing causes of behavior. This received version of Davidson's dogma is my primary target. Before moving on, however, I should note that Davidson's dogma, as received, does not perfectly represent Davidson's actual view. Two oft-overlooked subtleties of Davidson's actual account of belief attribution warrant mention.

First, Davidson was a dyed-in-the-wool Rylean about the central purpose of lay belief attribution: in light of a belief, "an action is revealed as coherent with certain traits, long- or short-termed, characteristic or not, of the agent, and the agent is shown in his role of Rational Animal" (1963: 8). However, Davidson warned that "it is an error to think that, because placing the action in a larger pattern explains it, therefore we now understand the sort of explanation involved" (1963: 10). Whereas Ryle averred that belief explanation proceeds simply by establishing that an action counterfactually depends on a belief, Davidson required an additional element: belief explanations must also highlight an inner cause that produced the action under scrutiny.

Second, as argued in Chapter 1, Davidson did not ascribe to the view that the folk literally construe beliefs, themselves, as inner causes. Like Ryle, he held that "attitudes [including beliefs] are dispositions to behave in certain ways" (Davidson 1997: 72), and that the folk veridically attribute them as such (1991: 215). In considering objections to his dogma, Davidson conceded that beliefs, being dispositional states as opposed to events, could not be producing causes. Instead, belief attribution "gives a

cause only on the assumption that there was also a preceding event" (1963: 12); the folk construe beliefs as "causal conditions" (1974: 232) that are "very closely associated" (1963: 12) with inner causes.[79] Strictly speaking, the folk do not construe beliefs as causes. Instead, belief explanations obliquely refer to the "onslaught" of a belief—a mental event such as noticing, remembering, or changing one's mind—as the inner cause of the behavior being explained (1963: 12; 1970: 208). For example, by attributing the belief that the Earth is flat to Kyrie, I might be pointing towards the event of his feeling of skepticism as the inner cause that produced his behavior. Thus, the nuanced version of Davidson's dogma—the version actually endorsed by Davidson—is that the folk attribute beliefs partly in order to indirectly invoke mental events that are producing causes.[80]

To repeat: my central goal in this chapter is to reveal the lack of justification for the received version of Davidson's dogma defined above, rather than the nuanced version endorsed by Davidson himself. Nevertheless, my critique generalizes to Davidson's actual view; just as there is neither evidence nor a sound argument that the folk construe beliefs as inner causes, there is neither evidence nor a sound argument that belief attributions successfully explain behavior only by obliquely invoking a mental

---

[79] Davidson did not unpack this aspect of his view, but Fred Dretske (1988) has provided an influential account of how beliefs might serve as causal conditions—"structuring causes"—in explanations of behavior.

[80] Nevertheless, for the sake of expediency (Davidson 1974: 232), Davidson frequently lapsed into referring to beliefs themselves as causes, rather than causal conditions.

event that produced the behavior. Thus, following Davidson's own habit, I will usually write as if there is no difference between the two views.[81]

3. The folk craft of belief attribution

Fodor is right: Davidson's dogma is intuitively appealing. Why did (ecosocialist grad student) Patrick walk to the minifridge to fetch a bottle of meal replacement beverage Soylent 2.0?[82] Because Patrick believed there was a bottle of Soylent in the fridge. How does this belief explain Patrick's behavior? Well, it would explain Patrick's behavior if it played a role in producing it.

Intuitive appeal aside, the question at hand is whether we have objective grounds for inferring that this is actually how the folk construe the beliefs they attribute to each other in everyday life. One candidate ground is Davidson's argument, which I will assess in Section 6. Before getting there, I will explore sources of evidence that were unavailable to Davidson; over the last forty years, psychologists have conducted a flurry

---

[81] I am not denying that belief attributions ever serve to explain behavior by indirectly invoking mental events that produced the behavior in question, much less that belief attributions do sometimes indirectly invoke such mental events. On the contrary, both claims are compatible with accepting Ryleanism and rejecting Davidson's causalism. Davidson asserted that belief attributions *can only* serve to uniquely explain behavior by dint of invoking inner causes. My Rylean view is (as Davidson admitted) that the folk do not construe beliefs themselves as inner causes, and (contra Davidson) that belief attributions often, and primarily, serve to uniquely explain behavior without invoking inner causes. This view is compatible with the hypothesis that belief attributions sometimes offer a secondary explanation of behavior by obliquely invoking mental events as inner causes.

[82] Soylent 2.0, the ready-to-drink product from Rosa Lab's line of beverages designed to meet the nutritional requirements of the average adult, is popular among graduate students because of its ease-of-use and low price point. It tastes, and is textured, like watery pancake batter.

of empirical research on the folk craft of belief attribution. In this section, I will lay out current thinking about the cognitive mechanisms behind human belief attribution capacities and describe the model-theoretic account I prefer. In Sections 4 and 5, I will assess whether the two leading research traditions investigating human belief attribution capacities—mindreading research and attribution theory—have provided evidence for Davidson's dogma.

3.1. Cognitive scientific theories of belief attribution

In the 1980s and 1990s, two theories of mindreading held sway. Theory theorists (Premack & Woodruff 1978; Churchland 1981; Fodor 1987; Leslie 1987; Gopnik & Wellman 1992; Baron-Cohen 1995; Gopnik & Meltzoff 1997) argued that people construct and apply their belief attribution abilities in the manner of scientists, incorporating empirical evidence to build a predictive and explanatory theory of how beliefs relate to behavior. Simulation theorists (Gordon 1986; Heal 1986; Goldman 1989; Harris 1992; Gallese & Goldman 1998) argued that belief attributors are not scientific theorists, but empaths who imaginatively put themselves into believers' shoes. According to simulation theory, people attribute beliefs by pretending to be the believers who are their targets, and then introspectively observing the relationships between beliefs and behaviors that emerge in their simulations.

By the 2000s, hybrid theories of mindreading emerged, as theory theorists

incorporated elements of simulation theory (Perner 1996; Nichols & Stich 2003) and vice versa (Goldman 2006). As theory theory and simulation theory have matured, philosophers have also begun to propose novel theories of mindreading that fit neither traditional mold. For example, social narrativity theorists (Ratcliffe 2007; Hutto 2008; Bohl 2015) endorse simulation theorists' insight that belief attribution is embedded in non-theoretical social cognitive practices but argue the practice in question centrally comprise the construction and dissemination of stories understood as public artifacts (rather than private simulations). Meanwhile, model theorists endorse theory theorists' insight that belief attribution aims to (incorporate empirical evidence in order to) represent how beliefs relate to other mental states and behaviors but argue that belief attribution capacities more closely resemble scientific modeling than scientific theorizing (Maibom 2003, 2007, 2009; Godfrey-Smith 2004, 2005).[83]

Together, narrativity theory and model theory have rightly shifted the discussion towards the heterogenous, diachronic uses to which people put their belief attribution capacities. Whereas traditional theory theory and simulation theory focused on explaining how belief attributors predict and explain behavior, these new approaches

[83] In the last few years, there has also been a surge of interest in the theory that people directly perceive others' mental states (Gallagher 2008; Zahavi 2011; Smith 2013; Herschbach 2015; Helton forthcoming). However, whatever one makes of the possibility of direct social perception, it is nearly universally agreed that beliefs are not good candidate targets for being directly perceived (Spaulding 2015; but see Gallagher & Hutto 2008). As discussed in Chapter 1, Ryleans eschew their mislabeling as philosophical behaviorists and countenance unobservable aspects of belief.

highlight belief attribution's important rationalizing and regulative functions.[84] Mirroring the previous decade's convergence of theory theory and simulation theory, narrativity theorists have countenanced the idea that narratives facilitate working models of belief (Hutto 2009), and model theorists have recognized the key role that public story-telling plays in the folk practice of model building (Maibom 2009). Folk psychological pluralism (Morton 2007; Andrews 2012; Schaafsma et al 2015) has won the day; no one-dimensional theory captures the variety of social cognitive capacities that, taken together, compose the folk theory of mind.

3.2. A Rylean model theory

At least when it comes to belief attribution, I favor a model theory informed by narrativity theory, with an emphasis on the tenet that folk psychological models can be genuinely explanatory without positing productive causal relationships.

One major advantage model theory has over traditional versions of both theory theory and simulation theory is that it neatly accounts for belief attributors' dual ability to think theoretically about the general conditions on possession of a given belief and think practically about what it would take for a particular believer to possess that belief in a particular context. Like scientists, belief attributors have both "an understanding of a general structure or schematic pattern that can have many specific instantiations"

[84] These functions of belief attribution are discussed in Section 5, as well as Chapters 4 and 7.

(Godfrey-Smith 2005: 4) and "the ability to construct *specific* hypothetical systems to deal with particular empirical cases" (5). By wielding this dual ability, "folk-psychological attributors can rapidly put together specific, filled-out psychological profiles, to explain and predict the actions of individual agents" (6). Narratives about people's mental lives inform the construction of both types of model: they serve to deepen and broaden general-purpose models by revealing the diverse range of ways in which people can take the world to be a certain way, while also helping attributors understand the particular "contributions made by a person's character, history, and larger projects" to her style of belief (Hutto 2009: 14). Through locating narratives as one of many sources of enrichment for both general-purpose and believer-specific models of belief, model theory uniquely reveals "how we manage to systematize an extraordinary range of phenomena and understand them as different manifestations of the same general principles" (Maibom 2009: 374–375).[85]

Model theory also usefully contrasts the *construction* of (and general facility with) both general-purpose and believer-specific models with the *construal* of those models. Belief attribution worthy of the name must construe beliefs as contentful, intentional, and truth evaluable. Beyond these basic characteristics of believing, however, attributors have considerable flexibility with regard to how they understand believers as taking the world to be certain ways. For example, both general-purpose and believer-specific

[85] In Chapter 6, I will provide a Rylean account of how both general-purpose models and believer-specific models figure in the metaphysical constitution of belief.

models of belief situate beliefs relative to other mental states—including other attitudes as well as thoughts and feelings—and behaviors. But models of belief do not, on their own, dictate the nature of these relationships. A number of possible construals are available to belief attributors. As Peter Godfrey-Smith notes, these construals track positions occupied by metaphysicians of belief:

> Roughly, they fall on a scale from more instrumentalist to more realist construals. The model can be seen merely as an input-output device. The model can also be seen as a rough and minimal map of inner causal structure; its basic elements and the interactions between them can be seen as corresponding very roughly to real psychological structures. The model can also be seen as a much more detailed map of how psychological processes work. Fewer features or more features of the model can be seen as having distinct, well-demarcated analogues in target systems. All these construals are *available* but none is *mandatory*. The folk-psychological model does not dictate its own construal. If we ask "What is folk psychology itself committed to?", the answer is "Nothing." (Godfrey-Smith 2005: 9–10)

The guiding question of the present chapter is: which construals of models of belief do folks actually employ, in their everyday belief attribution practices? Folk psychology

itself may be committed to nothing. But what are folk belief attributors implicitly committed to, by virtue of how they ordinarily construe the beliefs they model when engaged in the folk craft of belief attribution? Davidson's dogma furnishes this question with an answer that supports the conflation of attitudes of belief with cognitive states of belief; Ryle furnished it with an answer equipped to resist that conflation.

Godfrey-Smith allows that "everyday folk-psychological interpretation displays a good deal of flexibility with respect to which kinds of understanding it is used to achieve," but insists, in line with Davidson's dogma, that lay belief attribution is often "used to try used to try to make deep claims about inner causal processes (whether this activity is justified or not)" (2005: 10). There is no doubt that models of belief *can* be so construed; Fodor (1975; 1987) so construes them when positing a cognitive architecture that obeys the inner logic of folk psychology. As I will endeavor to show, however, there is every reason to doubt that the folk construe beliefs as inner causes when not explicitly theorizing about cognitive science.

Putting the debate over construals momentarily aside, the biggest challenge facing model theorists lies in explaining how models of belief are functionally implemented in empirically well-attested cognitive mechanisms. The functional mechanisms underlying belief attribution are currently by-and-large a matter of speculation. Nevertheless, there are good reasons to adopt the following speculative, high-level functional sketch of how adult humans model each others' beliefs.

Models of belief, like stereotypes, are a species of *characterization* (Camp 2015, forthcoming): they are associative grasps on idealized patterns of properties. I possess a model of the belief that there is Soylent in the fridge insofar as I recognize that people with that belief will be inclined towards certain Soylent- and fridge-directed thoughts, feelings and behaviors. Like stereotypes (Amodio 2014), general-purpose models of belief are rapidly and efficiently accessible. If someone says "quarterback", you automatically imagine a square jaw and can-do demeanor (Camp 2015); if someone says "Flat Earther", you automatically imagine a distrust of globes and tendency to spend time on internet message boards.

Taking these characterizations and applying them to particular believers is neither automatic nor cognitively cheap (Apperly et al. 2006), although the cognitive cost varies with context (Apperly 2011: Chapter 5). Some contexts—attributing beliefs to strangers, or trying to distinguish the believers from the non-believers—simply require attributors to measure people up to stored general-purpose models of belief. Other contexts—attributing beliefs to loved ones, or nuanced protagonists of novels—lead attributors to construct specialized models of belief, on the basis of general-purpose models, other general characterizations, and characterizations specific to the believer in question. For example, my characterization of Kyrie Irving's wholehearted embrace of his own celebrity—and attendant world-stage—significantly alters the pattern of behaviors, thoughts, and feelings that are associated with his particular brand of Flat

Eartherism in my model of that belief.[86]

All in all, belief attribution capacities center on the construction, storage, retrieval, ongoing reconstruction, and application of models of belief. In a typical case, when somebody undertakes to attribute a belief, they retrieve general-purpose and/or relevant specialized models of that belief, condition them with information derived from stereotypes and/or believer-specific characterizations (including models of character traits, moods, and other relatively stable aspects of mind), and perhaps generate a new believer-specific model. They then infer that the target believer fits (or does not fit) the model they have brought to bear, given prior knowledge, behavioral evidence, and context clues.[87]

This rough order of operations is a substantive and empirically testable—and reasonably empirically well-attested (Simons & Keil 1995; Wellman & Gelman 1998; Wellman 2014: 141–142)—aspect of this model-theoretic account of belief attribution. Abstract (general-purpose and specialized) models of belief are cognitively prior to particular attributions of belief to particular believers. This cognitive priority is to be expected insofar as narrativity theorists are right that attributors construct models of belief largely on the basis of conversations—including narratives and explicit parental

---

[86] I will have much more to say about believer-specific models of belief in Chapter 6.

[87] Wellman (2014) and Westra (2018a, 2018b) flesh out largely similar sketches of the functional details of belief attribution capacities within a hierarchical predictive coding (Clark 2016) account of cognition. I am sympathetic with this approach, but do not want to commit to any particular cognitive architecture in these pages.

instruction (Hutto 2008: Chapter 7; Wellman 2014: Chapter 2)—rather than first-hand proto-scientific experimentation. Even in cases where attributors have constructed a new believer-specific model based on observable evidence, they do not necessarily attribute belief. (I have a clear working model of how Patrick *would* live *if* he believed the Soylent is in the cupboard rather than the fridge, and of how Kyrie would live if he believed the Earth is round.)

4. The (lack of) evidence from mindreading research

Again, these functional details of human belief attribution capacities are admittedly speculative. What, then, has been the subject matter of the previously mentioned forty-year flurry of empirical evidence? By and large, experiments on human mindreading have concerned (not the functional implementation but) the ontogenetic development of belief attribution capacities.[88]

4.1. The empirical literature on lay belief attribution

Around the age of four, most humans explicitly demonstrate the ability to differentiate between their own true belief and the false belief of another agent (Wimmer & Perner 1983) or their own past self (Hogrefe, Wimmer, & Perner 1986; Gopnik &

[88] There is also a gigantic literature on the neural localization of belief attribution capacities, which I cite when appropriate, but this cognitive neuroscientific work does not reveal much about cognitive mechanisms.

Astington 1988). This developmental timeline fluctuates from culture to culture, but the central result is cross-culturally robust (Slaughter & Perez-Zapata 2015). At some point in their cognitive development, all normally abled humans gain the ability to attribute (true and false) beliefs to themselves and other agents. For many years, mindreading research focused almost exclusively on whether young children were able to pass various false belief tests. This "neurotic task fixation" (Gopnik et al. 1994) was well-motivated: more than any other developmental milestone, the ability to model false beliefs—the understanding that people can sincerely misrepresent states of affair—unlocks children's social worlds.

Since the turn of the century, however, researchers have begun to test mindreading more systematically. A barrage of evidence suggests that children much younger than four—even within the first year of life—demonstrate some capacity for mindreading, though they lack the ability to reason explicitly about false beliefs (Scott & Baillargeon 2017). Meanwhile, research on adult mindreading reveals that the ability to track behavioral cues associated with what somebody believes, the ability to infer what somebody believes, the ability to hold in mind what somebody believes, and the abilities to predict, explain, and interpret somebody's behavior on the basis of what they believe all make distinct (and often significant) demands on the cognitive resources of mindreaders (Apperly 2011). Experiments also suggest that the explicit attribution of (especially false) belief is subject to egocentric biases, even when executed by skilled

adult mindreaders (Epley et al. 2004; Birch & Bloom 2007).

These findings on the belief attribution capacities of humans at every stage of development have led some prominent researchers to adopt two-systems accounts of mindreading.[89] According to two-systems accounts, the high-level ability to model and reason about beliefs does not come online until the age of four. However, low-level mindreading (or submindreading) abilities allow infants to track beliefs in an automatic and cognitively efficient manner. Low-level mindreading may consist in modular mechanisms specialized for tracking mental attitudes (Butterfill & Apperly 2013a; Roessler & Perner 2013), or it may recruit domain-general cognitive mechanisms, "such as those involved in automatic attentional orienting and spatial coding of stimuli and responses" (Heyes 2014: 140). In either case, low-level processes allow mindreaders to track behavioral cues associated with belief, and respond appropriately, without actually attributing beliefs. As Cecilia Heyes puts it, the low-level mindreader "is doing the things that he or she would do if he or she was [attributing belief], but the cognitive processes controlling his or her behavior do not represent mental states" (2014: 132).[90]

---

[89] Nativists have pushed back on the two-systems interpretation of the evidence (Carruthers 2015b; Westra 2016; Scott & Baillargeon 2017), but a Rylean account of mindreading jibes with one-system accounts as well. According to one-system Ryleanisms, mindreaders' ability to attribute beliefs (qua patterns of dispositions) is of a piece with their ability to track dispositions. Nativist, as opposed to developmentalist (Wellman 2014), one-system accounts make the further claim that this univocal ability is innate. None of the evidence that leads nativists to their nativism entails that infants must construe beliefs as inner causes.

[90] Butterfill and Apperly (2013a: §4.3) disagree with Heyes on this point, taking low-level mindreading to involve the attribution of minimal mental states which they term 'registrations'. This might appear to be a stark difference: it might seem like Butterfill and Apperly take infants to attribute inner causes, whereas Heyes does not. After all, Butterfill and Apperly assert that

Together with "the rich endowment of social knowledge that we gain through development" (Apperly 2011: 155; cf. Ratcliffe 2006; Andrews 2012), low-level mindreading abilities lay the cognitive foundation for the high-level ability to represent beliefs in a controlled but cognitively inefficient manner.

On this two-systems account, the ability to attribute belief co-opts low-level capacities to track how believers register their environments and are disposed to think, feel, and behave in context. The robust predictive success of these low-level capacities is made "possible because our everyday social lives include a high degree of regularity that can be exploited by inflexible … processes, and captured in social scripts, schemas, roles and habits" (Apperly 2011: 155). Hijacking the information from low-level capacities to model and explicitly attribute beliefs is just one more way to pick out

---

"registrations are intermediate variables and play a subset of the causal roles characteristic of belief. For registrations, like beliefs, … causally influence actions" (2013a: 621). However, they strictly limit the kind of causal influence that registrations have on actions: whereas "beliefs interact with desires in influencing the means we select to achieve a goal, [infants] may think of [registrations] as fixing parameters on basic object-directed actions (e.g., where someone will reach or walk to)" (Apperly & Butterfill 2009: 957); "registration differs from belief in … not interacting in arbitrarily complex ways with other states, and in the simplicity of its parameter-setting role in causing actions" (Butterfill & Apperly 2013a: 629). In other words, they deny that the folk construe registrations as cognitive mechanisms; instead, the folk construe registrations as relations agents bear to situations, which constrain how those agents might act. Moreover, in response to Zawidzki (2013), Butterfill & Apperly (2013b: §5) make clear that, "inspired by Woodward (2003)" (2013b: 12), they are working with dependence—rather than production—notions of 'intermediate variables' and 'causal influence' when describing registrations (see endnote 8). In sum, Butterfill and Apperly take low-level mindreaders to attribute minimal mental states, but do not consider these states to be inner causes in the sense at stake. Heyes (2014: 140) and Roessler and Perner (2013: 36–37; Perner & Roessler 2009: 207–209) agree with Butterfill and Apperly that low-level mindreading fixes on dependence causes, though they disagree that the dependence causes in question should count as mental states.

patterns in mindreaders' noisy social environments. However, as Roessler and Perner argue, the attribution of belief rises above other forms of social knowledge in that it "enables [the belief attributor] to see a person's reasons relative to a different perspective" (2013: 46): the perspective of the believer herself. Whereas social scripts help low-level mindreaders understand reasons why an agent *might* act, belief modeling and attribution enable an understanding of the reasons an agent actually possesses.

What is involved in the possession of a belief? At least this: to possess a belief is to represent—or misrepresent—the world as being some way. According to Josef Perner's (1990) influential interpretation, mindreaders who pass false belief tasks demonstrate their understanding that believers act in accordance with subjective (mis)representations of their environments, rather than objective features of those environments themselves. Younger "children fail to understand belief because they have difficulty understanding that something represents; that is, they cannot *represent* that something is a *representation*" (Perner 1990: 186). On Perner's view, children have no problem realizing that people act for reasons. The tricky aspect of belief attribution is that "it presents its practitioners with a *perspective problem*" (Roessler & Perner 2013: 40): kids have to figure out that believers act in accordance with representations that comprise idiosyncratic perspectives.

Two-systems theorists stress that humans use explicit belief attribution sparsely, to supplement more routine low-level social cognition:

> In human adults, both systems exist in parallel. The cognitively efficient system plays a central role in guiding on-line social interaction and communication. The cognitively flexible system enables adults to engage in top-down guidance of social interaction (such as anticipating what the audience of a lecture might know, or working out how one misjudged the audience afterwards), and in explicit reasoning about the causes and justifications of mental states (as in everyday practical reasoning, or jurisprudence). (Apperly & Butterfill 2009: 966)

In most social situations, it is enough to use behavioral cues to prompt cognitive routines that automatically guide social interaction. Occasionally, though, mindreaders get a better handle on how to interact—or how they should have interacted—with their peers by explicitly reasoning about how they represent their shared world.

4.2. A Rylean interpretation of the evidence about lay belief attribution

The two-systems account of mindreading, thus stated, is compatible with Ryle's account of lay belief attribution. Low-level mindreading, combined with social knowledge, enables mindreaders to track and respond appropriately to believers' registrations and behavioral dispositions, and high-level mindreading enables mindreaders to group and make sense of salient patterns of those dispositions (along with dispositions to have associated thoughts and feelings) under the heading of 'belief'.

By attributing beliefs, mindreaders fit otherwise puzzling behaviors into coherent patterns, and thereby anticipate behaviors, thoughts, and feelings for which there are not immediately available behavioral (or social situational) cues.

As Perner argues, full-blown belief attributors find these patterns coherent because they grasp that believers represent the world from their own idiosyncratic perspective. On the Rylean interpretation, this representational aspect of belief ties the relevant dispositions together. What all of the dispositions that compose any given belief have in common is that they are dispositions towards thoughts, feelings, actions and reactions that make sense from the perspective of somebody who represents the world in a certain way. For example, Kyrie's tendencies to roll his eyes during a total solar eclipse and distrust science teachers make sense from the perspective of a free-thinking young man who represents the Earth as flat. To say that somebody believes *p* is not to say that there exists a discrete cog in the believer's cognitive system that carries the representational content *p*; instead, it is to say that there exists a believer who lives as if *p*. This Rylean interpretation accounts nicely for the logic of representation, including the notion that beliefs represent the world from a unique perspective.[91] All it dispenses

[91] This Rylean account of the representational content of belief is admittedly schematic. It has been developed in more detail by Robert Matthews (2007), building off of the work of Ruth Barcan Marcus (1990) and others; I also discuss it, still too briefly, in Chapter 5.2.7. According to Matthews's measurement-theoretic view of propositional attitude reports, belief attributions measure beliefs by relating believers to representational contents, but beliefs themselves are not contentful. In any case, every naturalistic account of the representational content of belief on offer is (infamously) less than satisfying, so the schematicity of current Rylean accounts of representation is not necessarily a strike in favor of Davidson's dogma.

with is Perner's conviction—presented without appeal to empirical evidence—that the folk construe representations as discrete inner states which play productive causal roles in the mind (1990: Chapter 2).[92] By merging with Ryleanism, the two-systems account of mindreading can distinguish belief attribution from less sophisticated forms of social cognition without assuming that belief attributors construe beliefs as inner causes.

4.3. Mindreading researchers assume Davidson's dogma

Nevertheless, mindreading researchers typically assume the truth of Davidson's dogma. On the prevailing two-systems view, the construal of beliefs as inner causes is part of what distinguishes high-level mindreading from low-level mindreading. Perner insists that, perhaps unlike low-level mindreading, belief attribution explains action "by being explicit about the intervening mind" (2010: 257), and Apperly and Butterfill (2009: 958) write that "whatever [infants] represent does not involve the causal and justificatory structure that is constitutive of adults' flexible belief reasoning (Davidson, 1989, 1995)." Infants do not (even implicitly) construe beliefs as inner, producing causes. Thus, if the folk construe beliefs as cogs in cognitive systems, then this construal is due

[92] There is good evidence (Estes, Wellman & Woolley 1989; Watson, Gelman & Wellman 1998) that, from an early age, people understand conscious thoughts and feelings—what Piaget termed 'mental things'—to be inner, unobservable states. However, even proponents of Davidson's dogma (Perner 1990: 170–178; Wellman 2014: 47) distinguish between beliefs (and other attitudes) and conscious mental things. To my knowledge, there is no corresponding evidence that people construe beliefs as inner. (And, as argued in Chapter 1, the existence—and folk recognition—of inner states of consciousness does not impugn Ryle's theory of mind.)

to the exercise of the later maturing ability to attribute belief explicitly.

To my knowledge, the mindreading literature provides no evidence—indeed, features no experiments designed to test the hypothesis—that explicit adult mindreading does construe beliefs as inner causes. It is, of course, consistent with the evidence that kindergarteners construe beliefs as inner causes, and pass false belief tasks by explicitly reasoning that the agent's false belief causally produced their behavior. Studies do show a significant correlation between the ability to pass false belief tasks and the ability to reason counterfactually (Riggs & Peterson 2000), and counterfactual reasoning is in turn linked to causal reasoning (Gerstenberg et al 2015). But these links are not enough to establish that the folk construe beliefs as inner causes. Recall Hall's distinction between counterfactual dependence causes and producing causes.

Belief attribution is linked to counterfactual reasoning, which is linked to causal reasoning; but counterfactual reasoning is not—indeed, logically could not be—more closely linked to reasoning about producing causes than reasoning about dependence causes (Menzies 2015). Counterfactual reasoning is, after all, the core mechanism of the folk understanding of dependence causation (see footnote 76). Thus, none of the available evidence speaks against the Rylean option that mature mindreaders construe behaviors as merely counterfactually dependent on beliefs without positing beliefs as inner causes.[93] The hypothesis that the folk construe beliefs mechanistically—as

[93] The Rylean U.T. Place (1996: 30) has argued from ordinary use and epistemic limitations that intentional behaviors are counterfactually dependent on beliefs, which are in turn

producing causes—and the hypothesis that the folk construe beliefs as mere counterfactual dependence causes are equally consistent with the evidence from mindreading research.[94]

Why, then, do mindreading researchers like Perner and Apperly assume that belief attributors construe beliefs as producing causes? Presumably because it is intuitive and, after all, the standard view (thanks to Davidson). Butterfill and Apperly (2013a: 611) assert that "on any standard view, propositional attitudes form complex causal structures … and are individuated by their causal and normative roles in explaining thoughts and actions (Davidson 1980, 1990)."[95] Apperly and Butterfill also recruit Davidson's dogma to help explain the difference between high-level mindreading and low-level mindreading.

Why is belief reasoning costly? At least part of the cost arises from the type of reason-giving explanation in which beliefs feature. For example:

---

counterfactually dependent on brain processes. For my purposes, the relevant issue is the folk construal of beliefs, rather than the metaphysics of dispositions. Whether or not ordinary use and epistemic limitations are arbitrators of fundamental metaphysics, they are indispensable sources of insight into folk construals. If the folk construe beliefs as identical to patterns of dispositions, and do not construe dispositions as identical to inner producers of behavior, then the folk do not construe beliefs as identical to producing causes.

[94] I will discuss research that does tell between these two hypotheses in Section 5.

[95] Perner and Roessler agree with Butterfill and Apperly that this standard view is "as Davidson has taught us" (Perner & Roessler 2009: 207; Roessler & Perner 2013: 26).

> She reached for the salt container because she saw the white grains, and—believing them to be sugar—intended to sweeten her pie.

> Reason-giving explanations like this one have several features (Davidson 1980, 1990). First, they involve complex causal structures: the perceptions influence beliefs which, together with desires, lead to intentions which guide action. Note how reason giving explanations invoke, explicitly or implicitly, multiple interacting causes, some of which may be far removed in time and space from the salient causes of an action. (Apperly & Butterfill 2009: 25; see also Butterfill & Apperly 2013a: 629)

Apperly and Butterfill marshal strong evidence in support of their hypothesis that explicit belief attribution is cognitive costly. They do not marshal any evidence, beyond citing Davidson, that part of this cost is due to the difficulty inherent in representing beliefs as inner causes that produce behavior.

This lack of evidence is striking. Nevertheless, the literature on mindreading may not be the best place to look for evidence to back up Davidson's dogma. Mindreading researchers investigate the cognitive ability to attribute belief, which is not, in and of itself, exactly what was at issue in the debate between Davidson and Ryle. The issue was the role that belief attribution plays in lay explanations of behavior.

5. The (lack of) evidence from attribution theory

Psychologists in the field known as 'attribution theory' have been studying how the folk explain behavior since the pioneering work of Fritz Heider in the 1940s and '50s.

5.1. The empirical literature about lay behavior explanation

According to classic Heiderian attribution theory, folk explanations of behavior divide in two: some explanations construe behavior as produced by 'situation' (or 'external') causes and some explanations construe behavior as produced by 'person' (or 'internal') causes (Kelley 1967; Ross 1977). This person/situation dichotomy is empirically motivated by phenomena such as the Fundamental Attribution Error—the tendency to overestimate person factors, and underestimate situation factors, when explaining the behavior of others (Jones & Harris 1967). Heiderian attribution theory generates a straightforward verdict on Davidson's dogma. The folk either construe beliefs as person causes or as situation causes, and they surely do not construe beliefs as situation causes. (Nor do the folk lump beliefs with more nebulous person causes, like personality traits or abilities.) Instead, Heiderian attribution theorists number beliefs among "the core [inner] processes which manifest themselves in overt behavior" (Heider 1958: 34).[96]

---

[96] A caveat: just as Davidson was not actually a proponent of Davidson's dogma, Malle (2004: Chapter 1, 2011: 301–304) persuasively argues that Heider was not actually a classic Heiderian attribution theorist.

Since the late 1970s, however, a strong minority voice has spoken against this way of carving up lay explanations of behavior (Buss 1978; Kalish 1998; Malle 2011). According to these dissidents, many folk explanations of behavior invoke neither situation causes nor person causes as typically understood. Instead, many explanations invoke reasons for acting. Instead of identifying what produced the behavior to be explained, reasons explanations identify what rational justification the agent had for acting as they did. When people demand an explanation for Patrick's behavior of walking to the fridge, they do not want to know what (or how) cognitive machinery produced his footsteps. They want to know what rationalized his purposeful journey to the fridge.

Bertram Malle has developed the most thorough new-wave attribution theory to date. Malle is struck that high-level mindreading, including the attribution of belief, "is anchored in the folk concept of intentionality," that is, the purposiveness of action. Malle allows that some behavior explanations—particularly explanations of accidental behaviors—make straightforward reference to classic inner person causes, "in the same mechanical way that physical causes explain physical events" (2004: 61). (Consider the explanation that 'the boy cried out because he felt a pain'.) Malle denies, however, that explanations of accidental behaviors have the same explanatory structure as explanations of intentional behaviors, such as 'the boy cried out because he believed doing so would get his mother's attention'. According to Malle, classic attribution theory

fails to account for systematic differences between explanations of intentional behaviors and explanations of unintentional behaviors with inner causes. Both of these varieties of explanation attribute the source of behavior to the person rather than the situation, but they serve strikingly different explanatory purposes.[97]

Malle's research has shown that these systematic differences between explanations of accidental behaviors and explanations of intentional behaviors can be put down to the former invoking mere inner causes whereas the latter invoke reasons—that is, mental state terms coded as beliefs, desires, or values. People are consistently good at categorizing behaviors as intentional or unintentional (Malle & Knobe 1997). And the more intentional people perceive a behavior to be, the more likely they are to invoke reasons to explain that behavior. Similarly, people rate behaviors as more intentional when they have been explained by invoking reasons, relative to when they have been explained by invoking mere inner causes (Malle 1999: 28–31). The take-away messages from these findings are that the folk attribute beliefs (and other reasons) to explain actions which agents intended, and that the folk conclude that agents must have intended behaviors that have been explained in terms of beliefs.

This fixation on explaining intentional actions in terms of reasons emerges early in human life. Henry Wellman's lab has shown that approximately 70% of young

---

[97] According to Malle's (2004) full theory, lay explanations of intentional behavior invoke reasons, causal histories of reasons, or factors that enable agents to act for their reasons. All three kinds of explanation invoke reasons; none invoke mere inner causes.

children's 'why-questions' request explanations for actions (Hickling & Wellman 2001; Frazier, Gelman & Wellman 2009), even though kids do not thoroughly grasp the range of ways in which beliefs explain actions until high-level mindreading comes online (Wellman, Cross & Watson 2001). When preschoolers answer why-questions themselves, they draw a clear line between reasons and causes, distinguishing between mistaken actions requiring explanation by reference to beliefs and accidental actions requiring explanation by reference to producing causes in the physical environment (Schult & Wellman 1997; Inagaki & Hatano 2002). In line with Malle's findings, Wellman's lab has found that whereas "preschoolers provided physical explanations almost exclusively for physically caused human movements," "nearly 100% of children's explanations for intended actions were psychological" (Wellman 2014: 45): that is, referred to attitudes like beliefs or desires. Moreover, researchers have shown that parent-child conversations regarding how actions causally depend on beliefs significantly influence how rapidly children develop the competence to pass false belief tasks (Dunn & Brown 1993; Bartsch & Wellman 1995: 137–142), whereas parent-child conversations regarding physical causality do not speed up mindreading development (Ruffman et al 2002).

In order to further examine the link between belief attribution and judgments of intentionality, Malle and his lab have investigated the motivations for belief attributions in lay explanations of behavior. Whereas explanations of unintentional behaviors focus

on answering the question 'what produced the behavior?', explanations of intentional behaviors focus on answering a more diverse range of questions. For example, belief explanations are used to answer questions such as 'why on Earth would the agent purposefully act in such a weird way?' (Korman & Malle 2016), 'what justified the agent in acting in this way?' (Malle et al 2000), 'should I blame the agent for having acted in this way?' (Monroe & Malle 2016), and the all-important 'how can I appear cool despite having acted in this way?' (Malle, Knobe & Nelson 2007). None of these questions can be answered adequately by bare reference to the producing cause of the agent's behavior, in its guise as a producing case rather than a reason. Instead, they must be met by an illustration of the agent's perspective which puts the anomalous, apparently irrational or immoral, or uncool behavior in context. Attributions of belief provide good answers to a wide range of questions about intentional behaviors because they reveal the agent's purpose—or at least a purpose the agent could have had—for acting a certain way (Malle 2004: 72–75).

From the age of six months, humans interpret intentional behaviors teleologically—that is, in terms of the purpose of the action—rather than mechanistically—in terms of how the action is produced (Csibra & Gergely 2013). Tania Lombrozo has argued that this distinction between mechanistic and teleological explanations is "psychologically real and cognitively deep" (2010: 329) by adducing evidence of a link between these two varieties of lay explanation and the two varieties of

causation distinguished above. Previous work had shown that the folk understand teleological explanations as a kind of causal explanation (Kelemen & DiYanni: 2005; Lombrozo & Carey: 2006), without differentiating between producing causes and dependence causes. Lombrozo hypothesized that "teleological explanations, by virtue of being 'mechanism-independent,' should encourage a criterion for causation in terms of the dependence of the effect on the cause," and that "in contrast, mechanistic explanations should encourage a criterion for causation that is more sensitive to aspects of the [productive] transmission from cause to effect" (Lombrozo: 2010: 325). Her experiments bore out these predictions.

Whereas people tend to judge causality by both production and dependence criteria when providing mechanistic explanations of accidental behavior, they tend to judge causality solely in terms of dependence when providing teleological explanations of intentional behavior.[98] In particular, "when an effect counterfactually depended on two agents who acted intentionally, participants provided very similar causal ratings, no matter that one agent … did not share a transference [i.e., producing cause] relationship with the effect" (325). In sum, when providing explanations of intentional behavior—including explanations in terms of belief—the folk show no clear sign of concern with productive causation. While "teleological explanations are causal explanations … they are nonetheless treated differently from mechanistic explanations:" they invoke reasons,

[98] Relatedly, Kominsky et al (2015) show that judgments of moral valence affect the counterfactual reasoning that underlies dependence explanations.

including beliefs, and they apparently do not invoke producing causes (Lombrozo & Vasilyeva 2017: 426). If the folk construe beliefs as inner causes that produce behavior, it makes no discernable difference to the explanations of intentional behavior that they give in terms of belief.

5.2. A Rylean interpretation of the evidence about lay behavior explanation

Like the two-systems account of mindreading, Malle, Wellman, and Lombrozo's findings concerning behavior explanation comport nicely with Ryle's account of belief attribution. From an early age, humans are obsessed with requesting and providing explanations that help them understand other people as purposive agents. By attributing beliefs, mindreaders make sense of each other in terms of patterns of dispositions to act, react, think, and feel as if the world were some way. Certain salient intentional behaviors are counterfactually dependent on these patterns. Mindreaders, recognizing this dependence relation, attribute beliefs in order to explain why an agent acted abnormally, asocially, or asininely. Why did Kyrie Irving sincerely assert that the Earth is flat? Well, he would sincerely say that only if he believes that the Earth is flat: only if he is the sort of person who goes around telling people not to believe everything they learn in school, watching conspiracy theory videos on YouTube, rolling his eyes when confronted with photographic evidence that the Earth is an oblate spheroid, squinting critically at the horizon, and so on. Attributing a belief to Kyrie reveals his (purported)

purpose in sincerely asserting the Earth is flat—that is how he takes the world to be—as borne out by his pattern of living. When the *Washington Post* explains his behavior by reporting that "Kyrie Irving believes the Earth is flat" (Russell 2017), the journalist shows no signs of construing Kyrie's belief as the inner cause that produced his utterance.

Given widespread agreement about the teleological function of folk explanations of intentional behavior, Davidson's dogma is a strictly stronger—less parsimonious—interpretation of belief attribution practices than Ryleanism. Researchers like Malle and Wellman readily acknowledge that belief explanations are different in kind than standard causal explanations: the folk attribute beliefs in order to rationalize and contextualize thoughts, feelings, and behaviors. Davidson's dogma adds to this baseline Ryleanism the proposition that the folk construe beliefs as inner causes. In short, Ryleans claim that beliefs are reasons; Davidsonians claim that beliefs are both reasons and causes. As discussed in this section and the last, there is no extant empirical evidence to justify this unparsimonious addition.

5.3. Attribution theorists assume Davidson's dogma

Nevertheless, Malle and Wellman assume that the journalist must construe Kyrie's belief in Flat Earth Theory as an inner cause of his utterances. Wellman cites Davidson when writing that belief "explanations can appear simple—'he thought Juliet

was dead'—but invoke complex causal structure: Perceptions influence beliefs, which together with desires lead to intentions that shape actions" (2014: 42), and notes in passing that one of the goals of belief explanation is "to identify the proximal cause of some event that has occurred" (46). Malle, meanwhile, asserts that "all folk explanations of behavior attempt to provide, among other things, an answer to a causal question—Why did this behavior occur, what brought it about?" (2004: 148). Although Malle takes great pains to divide mere inner causes of behavior from beliefs that rationalize behavior, he retains traditional attribution theorists' commitment to Davidson's dogma that intentional explanations invoke a "causal mechanism that is uniquely involved in producing intentional action—that of reasoning and choice" (2004: 61). Malle casts reasons (including beliefs) as the special species of inner cause that the folk construe as productive of intentional behaviors. But Malle does not base this commitment on the sort of compelling evidence that drove him to distinguish explanations that invoke reasons from explanations that invoke mere causes. His empirical research does not reveal that the folk construe beliefs as inner causes. Nor does he take it to. Instead, Malle writes that

> reasons are considered causally generative. As mental states, they are seen as bringing about the agent's decision to act. In that sense, people consider reasons

> to be 'causes' (Davidson 1963, Malle 1999), but causes with unique properties. (2011: 314)

In that 1999 article, Malle again cites Davidson: "philosophers have pointed out that reasons are themselves 'causes' if we define causes broadly as factors that generate an event such as behavior (Davidson 1963; cf. Locke & Pennington 1982)" (Malle 1999: 24). What about that Locke & Pennington reference? They provide an overview of the debate in action theory, before settling for "the current philosophical orthodoxy … that far from being incompatible, reasons are themselves but one kind of cause (Davidson 1963)" (Locke & Pennington 1982: 213). Like Malle, Locke & Pennington do not attempt to adduce empirical evidence in support of Davidson's dogma.

All in all, the evidence from mindreading research and attribution theory broadly substantiates Ryle's account of lay belief attribution, and lends no support to the addition of Davidson's dogma. Nevertheless, perhaps thinkers like Malle, Wellman, Perner, Apperly, Churchland, and Fodor are right to rest easy with their intuitions; perhaps Davidson demonstrated that the folk construe beliefs as inner causes way back in 1963.

6. Davidson vs. Ryle

Roessler and Perner (2013: 36) write that "as Davidson taught us, to explain why

someone got up at 3 a.m., it is not enough to assemble considerations—"justifying reasons"—that show this to have been the right thing for him to do. What is required is a causal explanation (Davidson 1963)." Without further ado, I will examine the details of Davidson's teachings.

6.1. Davidson's argument for his dogma

As confirmed by attribution theory, the folk take beliefs to rationalize intentional behaviors. Davidson agreed with Ryle on this point. But Davidson argued that when explaining intentional behaviors, it is not enough to nod to any old rationalizing belief. Believers have lots of beliefs that could serve to rationalize any given behavior. In most cases, only one of these is the primary reason upon which the believer actually acted. Only this primary reason properly explains the behavior in question.[99]

Imagine Patrick's search for Soylent. Patrick has all sorts of beliefs that might rationalize his behavior. As well as believing Soylent to be in the fridge, Patrick believes Soylent to be nutritious, thirst-quenching, filling, and delicious. These are all reasons that Patrick has for fetching Soylent from the fridge; when asked to rationalize Patrick's Soylent-fetching behavior, a mindreader might attribute any one (or more) of these beliefs. However, when asked to explain, rather than merely rationalize, Patrick's behavior, the mindreader's options are much more limited. They cannot attribute any

[99] Davidson held that a primary reason consists of a belief-desire pair, but that "it is generally otiose to mention both" (1963: 6). To avoid being otiose, I will generally mention only beliefs.

old reason Patrick might have had for fetching Soylent; they must attribute the reason that actually motivated him to fetch Soylent on the occasion in question. For example, it is only the belief that Soylent is filling that explains a hungry Patrick's Soylent-fetching behavior when, as a matter of fact, he fetches Soylent from the fridge because he believes Soylent to be filling.

By Davidson's lights, commonsense explanatory practices usually single out one belief-desire pair as the primary reason that uniquely explains an action. A laundry list of reasons that could rationalize Patrick's behavior cannot suffice when we ask why Patrick fetched Soylent. Only the primary reason that actually motivated Patrick suffices. How does Patrick's primary reason uniquely explain his behavior? What is the "mysterious connection" (1963: 11) between this reason, as opposed to other rationalizations for fetching Soylent, and the behavior it explains? Davidson countenanced only one satisfactory answer to this question: the primary reason that uniquely explains a behavior must be (or invoke) the producing cause of that behavior. "When we offer the fact of the desire and belief in explanation," he wrote, "we imply not only that the agent had the desire and belief, but that they were efficacious in producing the action" (Davidson 1974: 232). Thus, Davidson declared reasons to be causes.

Here is a reconstruction of Davidson's argument.

P1: There is one primary reason (or set of primary reasons) that uniquely

explains each action. (Davidson 1963: 3–8)

P2: If there is one primary reason (or set of primary reasons) that uniquely explains each action, then that reason (or set of reasons) must be (attributed to invoke) the inner cause that produced that action. (Davidson 1963: 9–12; Davidson 1974: 231–232)

C: The one primary reason (or set of primary reasons) that uniquely explains each action is (attributed to invoke) the inner cause that produced that action. (Davidson 1963: 12)

Davidson's argument is valid but unsound. P1 is dubious; in its place, I would propose that there is one reason (or set of reasons) that uniquely answers each question that seeks an explanation for an action, but that different explanation-seeking questions demand different reasons (Anscombe 1957; van Fraassen 1980; Schaffer 2007). Whether Patrick's belief that Soylent is nutritious or his belief that Soylent is in the fridge uniquely explains his behavior hinges on whether we ask 'why did Patrick walk to the *fridge* to fetch Soylent?' or 'why did Patrick walk to the fridge to fetch *Soylent*?'.

I will not linger on this point, for even if P1 is true, P2 is false. Construing reasons as inner causes is not the only way to account for the mysterious connection between an intentional behavior and the primary reason that uniquely explains that behavior. Action theorists have proposed several alternative accounts of the mysterious

connection over the years (Wilson 1989; Gordon 2002; Sehon 2005; Hutto 2011; D'Oro 2012; Tanney 2013).[100] I will not discuss any of them here. Instead, I will argue that Ryle's (1949) account of belief explanation, which numbered first amongst the accounts that Davidson's argument was designed to overturn (Davidson 1963: fn. 1), already had the resources to demystify the connection.

6.2. Ryle's anticipation of Davidson's argument

According to Ryle, belief explanations are context-placing explanations. A successful belief explanation places "an action in a context that renders it less puzzling" or outrageous (Tanney 2013: 164). In particular, a belief uniquely explains an action when it uniquely reveals how that action fits a broader pattern of living. When explaining somebody's actions by invoking his beliefs,

> the sense in which we 'explain' his actions is not that we infer to occult causes, but that we subsume under hypothetical and semi-hypothetical propositions. The explanation is not of the type 'the glass broke because a stone hit it', but more nearly of the different type 'the glass broke when the stone hit it, because it was brittle'. (Ryle 1949: 50)

[100] Other than Davidson himself, Alfred Mele (2003, 2017) has provided the most influential arguments in action theory for Davidson's dogma. See D'Oro & Sandis (2013) for a history of the debate.

In unpacking this analysis, Ryle foreshadowed Hall.

> There are at least two quite different senses in which an occurrence is said to be 'explained'; and there are correspondingly at least two quite different senses in which we ask 'why' it occurred and two quite different senses in which we say that it happened 'because' so and so was the case. The first sense is the causal sense. To ask why the glass broke is to ask what caused it to break, and we explain, in this sense, the fracture of the glass when we report that a stone hit it. The 'because' clause in the explanation reports an event, namely the event which stood to the fracture of the glass as cause to effect. (88)

The first sense of 'because' invokes a producing cause: the event of the stone striking the glass produced the event of the glass shattering.

> But very frequently we look for and get explanations of occurrences in another sense of 'explanation'. We ask why the glass shivered when struck by the stone and we get the answer that it was because the glass was brittle. Now 'brittle' is a dispositional adjective; that is to say, to describe the glass as brittle is to assert a general hypothetical proposition about the glass. So when we say that the glass

> broke when struck because it was brittle, the 'because' clause does not report a happening or a cause; it states a law-like proposition. (88–89)

The second sense of 'because' establishes a relation of counterfactual dependence between explanandum and explanans. For example, the explanation that 'the glass broke because it was brittle' "says, roughly, that the glass, *if* sharply struck or twisted, etc. *would* not dissolve or stretch or evaporate but fly into fragments" (89). Ryle noted that "people commonly say of explanations of this second kind that they give the 'reason' for the glass breaking when struck" (89). People invoke producing causes in explanations when they seek to identify what generated an event; people invoke reasons as explanations when they seek to understand an agent's (or object's) character and/or their purpose in acting.

Ryle went on to argue that belief explanations are "of the second type and not of the first type" (*ibid*); they invoke reasons, not producing causes. To attribute a belief

> is to say that [the agent] is inclined to do certain sorts of things, make certain sorts of plans, indulge in certain sorts of daydreams and also, of course, in certain situations to feel certain sorts of feelings. To say that he did something from that motive[101] is to say that this action, done in its particular circumstances,

[101] Ryle: "roughly, 'believe' is of the same family as motive words" (1949: 134).

> was just the sort of thing that that was an inclination to do. It is to say 'he *would* do that'. (92–93)

Ryle proposed that the folk construe beliefs as syndromes of dispositions to act, react, think and feel as if the world were some way. To believe Soylent is delicious is to be disposed to say "I love the taste of Soylent", think "mmm mmm mmm Soylent", imagine Soylent when hungry, fetch Soylent gluttonously, feel gustatorily satisfied upon drinking Soylent, and so on. Thus, Ryle suggested that when we offer explanations like "Patrick fetched Soylent because he believes Soylent is delicious", we do not mean that Patrick's belief produced his strides towards the fridge. Instead, we mean "Patrick, being the sort of dude who says 'I love the taste of Soylent, thinks 'mmm mmm mmm Soylent', and so on, is also the sort of dude who can reasonably be expected to fetch Soylent in this context."

At first glance, this explanation of Patrick's behavior might appear stunningly uninformative. However, this apparent lack of informativeness can be put down to the fact that we already know that Patrick is a Soylent-drinking sort of dude. We would not seek an explanation for Patrick's Soylent-fetching behavior in the first place. However, if we witnessed celebrity deep-frier Paula Deen walking to the fridge to fetch Soylent, then we might demand an explanation. The explanation that Deen is the sort of lady who talks favorably about, thinks fondly of, and walks fervently to the fridge to fetch Soylent

would be informative. As Ryle stressed, people use belief attribution to understand believers. Learning that Deen has a positively-valanced belief about the gustatory quality of Soylent is a revelation about her lived experience. (Compare the explanation that Deen believes Soylent is delicious with the explanation that Deen mistakenly believes the Soylent bottle to be filled with milk.)

Fast-forward sixty years. Patrick suffers from Alzheimer's. Rosa Labs (the manufacturer of Soylent) went out of business years ago. Patrick's granddaughter visits the nursing home, and observes Patrick rooting around in the fridge. She asks him what he is looking for; he answers that he wants to drink some Soylent. Patrick's granddaughter aptly explains his behavior by attributing the (false) belief that there is Soylent in the fridge. This explanation is informative because it contextualizes Patrick's odd behavior: he behaves as if Soylent is in the fridge *because* he believes there is Soylent in the fridge, and he believes there is Soylent in the fridge *because* he is in the grip of Alzheimer's disease. These 'becauses' give reasons: they place Patrick's behavior in the context of his belief, and place Patrick's belief in the context of his disease.

Davidson would not disagree. He allowed that

> when we ask why someone acted as he did, we want to be provided with an interpretation. His behavior seems strange, alien, outré, pointless, out of character, disconnected; or perhaps we cannot even recognize an action in it.

> When we learn his reason, we have an interpretation, a new description of what he did, which fits it into a familiar picture. (1963: 9–10)

So far so good. Davidson's worry was that Ryle failed to account for the mysterious connection between actions and the primary reasons that uniquely explain those actions. In Davidson's words, "we are left without an analysis of the 'because' in 'He did it because …', where we go on to name a reason" (1963: 11). Davidson insisted that the only viable interpretation of the 'because' in belief explanations implies that beliefs are (or invoke) inner causes that produce behavior.

How can a Rylean make sense of the mysterious connection, without understanding belief explanations to invoke inner causes? In true Rylean fashion: by denying that the connection is at all mysterious. According to Ryle, the 'because' in belief explanations is the 'because' of 'the glass broke when the stone hit it, because it was brittle', rather than the 'because' of 'the glass broke because a stone hit it'. In other words, rather than analyzing the 'because' as invoking producing causes, Ryle analyzed the 'because' as invoking patterns of dispositions, including the disposition to behave in the manner being explained. For Ryle, the so-called mysterious connection is just this: an intentional behavior is the manifestation of (one or more of) the dispositions that constitute the belief that explains that behavior.

Advocates of Davidson's dogma will not yet be satisfied. Patrick's behavior is a

manifestation of dispositions that partly constitute each of his beliefs that Soylent is delicious, nutritious, and filling. But when Patrick is hungry, only his belief that Soylent is filling actually motivates him to get off the couch. In this scenario, the mysterious connection is between Patrick fetching Soylent and his belief that Soylent is filling. If, with Ryle, we deny Davidson's dogma that the folk construe beliefs as inner causes, then it can seem mysterious why people construe Patrick's behavior to be a manifestation of his belief that Soylent is filling rather than a manifestation of his belief that Soylent is delicious.

However, this connection only seems mysterious because Patrick's behavior is underdescribed. The action to be explained by appeal to belief is Patrick *hungrily* walking to the fridge to fetch Soylent, as opposed to *gluttonously* walking to the fridge to fetch Soylent. Once described richly enough, Patrick's action emerges as the manifestation of a disposition that partly constitutes all and only those beliefs that uniquely explain his action. Indeed, the explanation 'Patrick fetched Soylent because he believes Soylent to be filling' functions by differentiating hungrily walking to the fridge from gluttonously walking to the fridge. These are distinct actions, even if they are "photographically and gramophonically as similar as you please" (Ryle 1949: 140). When seeking an explanation of an action in terms of a primary reason, people seek a richer understanding of the action itself, as well as an insight into the mental profile of the actor. On Ryle's view, the belief for which somebody acts is the pattern of thoughts,

feelings, actions, and reactions that uniquely rationalizes why *this person* would perform *this action*, under *this rich description*, in *this situation*.

Whether or not this captures how the folk actually construe beliefs, Ryle provided a viable analysis of the 'because' in belief explanations. The 'because' might point at a dispositional profile, rather than (directly or indirectly) towards an inner cause. Therefore, Davidson's second premise is false: beliefs can uniquely explain actions without being construed as causally producing them. Indeed, as we have seen, psychological evidence supports Ryle's analysis over Davidson's. Research on mindreading suggests that explicit belief attribution ties together the behavioral dispositions tracked by low-level mindreaders. Research on behavior explanation suggests that mindreaders attribute beliefs to provide reasons (as counterfactual dependence causes) for actions. None of this research suggests that the folk also construe beliefs as inner causes. The belief attributor seeks to understand what people are like, not as the cognitive neuroscientist seeks to understand the patient in the fMRI, but as the priest seeks to understand the parishioner in the pew.

7. Conclusion

I have argued that Davidson's influential argument for his dogma is unsound; in addition to construing beliefs as inner causes, there is at least one other principled way—the Rylean way—of accounting for the supposedly mysterious connection

between actions and the beliefs that explain those actions. I hasten to add, however, that I have not shown that Davidson's conclusion is false. The folk might obliquely invoke producing causes whenever they attribute beliefs to explain behavior. Even on Ryle's account of lay belief explanation, the attribution of belief encompasses the attribution of dispositions to have thoughts and feelings. The manifestation of those dispositions may constitute producing causes of actions and invoking these producing causes may be what renders belief explanations uniquely satisfactory. For example, Patrick's belief that Soylent is tasty is partly constituted by his disposition to feel an urge to taste Soylent. In order to successfully explain Patrick's Soylent-fetching behavior, mindreaders might have to obliquely invoke the manifestation of this urge as a producing cause. In other words, either Davidson's nuanced view or even Davidson's dogma might be true, even though his argument is unsound.

Nevertheless, I leave the reader with three facts. First, the debate in action theory is vexed and ongoing (the Davidsonian assumptions of most philosophers of mind and psychologists notwithstanding). Second, mindreading research and attribution theory have, to date, provided no empirical evidence that the folk construe beliefs as inner causes. Third, evidence provided by psychologists such as Apperly, Malle, Wellman, and Lombrozo accords well with Ryleanism, and less well with Davidson's dogma, given that the latter posits an additional, unsubstantiated dimension to the folk construal of belief.

On Ryle's view, people attribute beliefs every day in order to provide teleological explanations of actions and insights into the personalities of actors, not in order to provide mechanistic explanations of behaviors. The evidence adduced in this chapter forms an empirical framework in which Ryle's hypothesis is well-founded. I look forward to more research by psychologists on how people model beliefs—and especially how they construe those models. In the meantime, together with the lack of evidence for Davidson's dogma, conceptual considerations have pushed me to adopt a working distinction between attitudes of belief and cognitive states of belief. Chapter 4 will lay out these conceptual considerations.

Chapter 4: How beliefs are like colors

1. Recap and introduction

In Chapter 1, I argued that all interesting alternatives to paramechanical theories of belief—that is, alternatives to standard mind-brain identity theories, functionalisms, and eliminativisms—are versions of Ryleanism. In Chapter 2, I argued that the question of whether to collapse or support the distinction between attitudes of belief and cognitive states of belief lies at the heart of the debate between paramechanists and Ryleans. In Chapter 3, I argued that the primary paramechanical presumption behind the collapse of the distinction between attitudes and cognitive states—the dogma that the folk construe beliefs as inner causes—is backed up by neither empirical evidence nor sound philosophical argumentation. I have not yet offered positive arguments for Ryleanism, beyond the glimmer of support provided by the experiments canvassed in Chapter 3. I will begin to pick up that slack here in Chapter 4, by way of a cascading series of considerations in favor of distinctions between attitudes of belief and cognitive states pitched at neurophysiological, subpersonal, personal, and etiological levels of psychological explanation.

I will begin by arguing against standard paramechanical suggestions about how to identify attitudes of belief with cognitive states of belief, by way of an analogy between belief and color. Colors are neither illusory nor intrinsic physical or dispositional properties of objects. Instead, colors are properties objects have in virtue of

occupying the environments of color perceivers. Likewise, Ryleans hold that attitudes of belief are neither figments of the folk psychological imagination nor intrinsic physical or functional properties of objects. Instead, attitudes of belief are properties organisms have in virtue of inhabiting the environments of belief attributors. As such, they ought not be conflated with cognitive states of belief: properties organisms have in virtue of possessing cognitive systems with certain architectures. (Depending on one's theory of cognitive states of belief, this may mean having certain intrinsic physical properties, bearing certain psychofunctional relations to concrete mental representations, functioning in certain ways on a personal level, or having properties that play certain irreducibly normative evolutionary purposes.)

I will go on to argue that insofar as teleofunctionalism helps theorists get a grip on the nature of belief, it should be recruited to support (rather than collapse) the distinction between attitudes and cognitive states. Like colors, attitudes of belief evolved to be ecological signifiers, not cogs in cognitive systems. There would be no believers—no creatures properly characterized as having attitudes of belief—without belief attributors—creatures who characterize creatures as having attitudes of belief. The arguments behind this core interpretivist commitment provide positive reasons, not just to doubt Davidson's dogma, but to affirm the distinction between attitudes of belief and cognitive states of belief. The concluding section will summarize the master argument of Part I, which leverages the distinction between attitudes and cognitive states to establish minimal Ryleanism about belief.

2. Color

Reflection on the metaphysics of color provides an impetus to recognize that attitudes of belief and cognitive states of belief ought to be pulled apart.

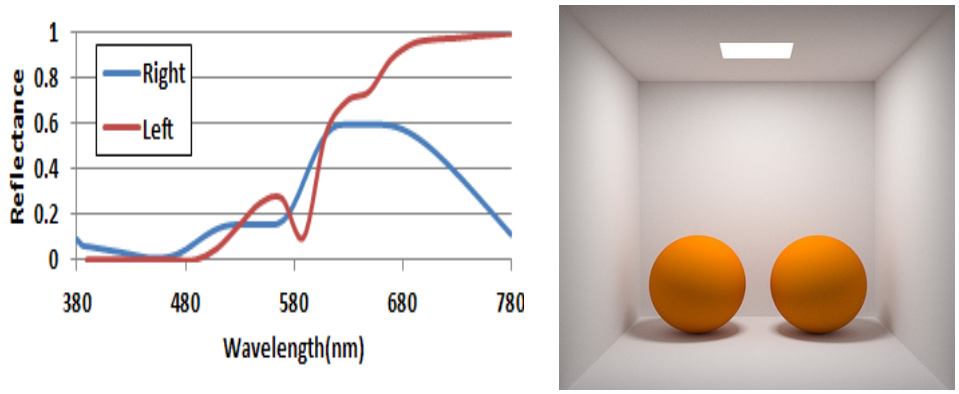
2.1. Against conflating perceivable colors with SSRs

To see why, consider a well-known problem for physicalism about perceivable color. Physicalism about color identifies perceivable colors with intrinsic properties of colored objects (Smart 1961; Armstrong 1968; Hilbert 1987), just as paramechanists identify attitudes of belief with cognitive states of belief. In particular, most physicalists identify colors with either surface spectral reflectances (henceforth SSRs) or the intrinsic physical properties of objects that realize SSRs. Surfaces absorb some light at each wavelength of the visible spectrum and reflect the rest. To have a particular SSR is to be disposed to reflect a particular percentage of light at each wavelength on the spectrum. Physicalists declare, rightly, that SSRs are the nonrelational dispositional properties of objects most closely associated with their colors. (For example, red objects are generally disposed to reflect more light in the longer wavelengths than blue objects.) By extension, the physical properties underlying SSRs are the intrinsic physical properties of objects most closely associated with their colors. Nevertheless, identifying colors with SSRs (or their physical realizers) is problematic because there is a double dissociation between the SSRs of objects and the perceived colors of those same objects.

First, the perceived colors of objects do not perfectly track their SSRs. Examples

of metamerism demonstrate that objects with strikingly different SSRs often appear identically colored to the same perceiver.

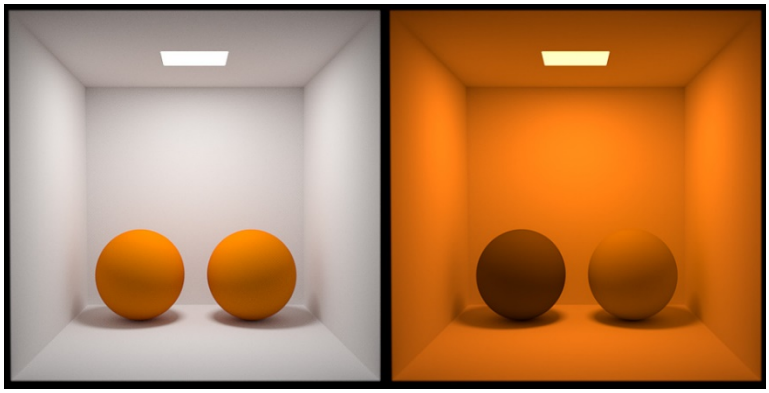
FIGURE 3: METAMERISM



A case of metamerism. The red and blue lines represent distinct SSRs that produce the same orange hue in white light for normal human perceivers.

Now, metamers only have identical perceived colors in some lights. Seen in another light, they exhibit different shades of orange.

FIGURE 4: METAMERS VIEWED IN ANOTHER LIGHT

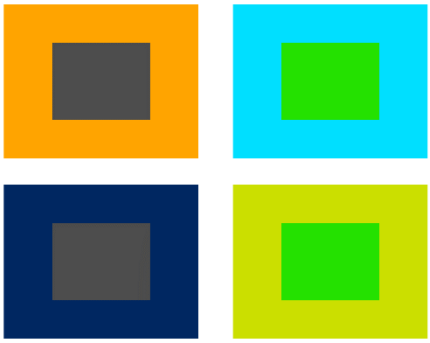


Metamers in white light (left) also viewed in orange light (right).

Nevertheless, in cases of metamerism, there is no principled way to settle on one of the SSRs as veridically represented by the shared perceived color. Likewise, there is no good reason to suppose that only one of the objects is *really* the color it appears to be. (Why would the orange light condition in the right panel of Figure 4 be decisive about the true color of the balls, as opposed to aberrant? After all, orange light illuminates far fewer human environments than white light.) Both objects in Figure 3 exhibit the same particular shade of orange for a particular perceiver in a particular context, despite their divergent physical properties. Thus, the perceived color of an object does not perfectly track its SSR; on the contrary, perceived colors are multiply realized by SSRs.

Nor does the SSR of an object fix its perceived color. Examples of intrasubjective perceptual variation demonstrate that objects with identical SSRs can appear differently colored in different contexts, even when perceived simultaneously by a single perceiver.

FIGURE 5: INTRASUBJECTIVE PERCEPTUAL VARIATION



A case of perceptual variation. Because of the colors of the outer rectangles, the inner squares that share an identical SSR are perceived as subtly differently colored. Image created by J. Taylor.

Moreover, the perceived colors of objects vary between perceivers, even when the context is held constant. A particularly stark—if atypical—example of this phenomenon recently became viral on the internet.

FIGURE 6: THE DRESS

The viral image of the dress: a case of intersubjective variation in color perception. Some people see the dress as black and blue; others see white and gold.

About 60% of people see the dress in Figure 6 as white and gold; about 30% see it as blue and black. Only about 1% of people report being able to (intentionally or unintentionally) switch between the two commonly perceived color configurations (Wallisch 2017: 4). Now, this stark difference in how the world is colored for different human perceivers rests on a knife's edge. Consider Figure 7.[102]

[102] Recent empirical research shows that subjects' expectations about illumination conditions—and attendant modulation of color constancy mechanisms—drive the differential perceived colors the dress exhibits, given ambiguous illuminance conditions, in Figure 6 (Wallisch 2017). In Figure 7, there is not much ambiguity in illuminance, so there is not much ambiguity in perceived color. Other researchers have offered (less convincing) explanations of the phenomenon in terms of differential macular pigment optical density (Rabin et al. 2016) or the top-down influence of knowledge of the color of the dress in normal illuminance conditions (Schlaffke 2015).

FIGURE 7: THE DRESS PHOTOGRAPHED IN DAYLIGHT

The dress in normal light: a case of widespread, approximate—though not absolute—intersubjective uniformity in color perception.

In most viewing conditions, the image of the same dress looks black and blue to nearly all humans (including those who saw it as white and gold in the previous image). Importantly, though, there remain subtle differences in the exact shades of black and blue experienced by different individuals (Webster 2015). Due to normal variation in human perceptual systems, I likely see the dress in Figure 7 as either a slightly redder or slightly greener shade of blue than you. Whether the dress's blue color is *really* redder or greener is not settled by reference to the dress's SSR.

It is extremely likely that there are more dramatic differences in how the perceived colors of objects vary between perceivers of different species. Animals with only one or two types of cone cell almost certainly see a differently colored world than trichromatic humans. The same goes for many fellow trichromats: red mason bees, for instance, detect shorter-wavelength 'ultraviolet' light reflected off of surfaces but not longer-wavelength 'red' light, whereas humans detect 'red' light but not 'ultraviolet'

light (Menzel et al. 1988).

FIGURE 8: INTERSPECIFIC VARIATION IN COLOR VISION

On the left: a crocus as seen by a human. On the right: how crocuses might look to bees.

Needless to say, the crocus pictured in Figure 8 has the same SSR for bees as for humans; nevertheless, it very likely exhibits unmistakably different colors for the two species of trichromat.

These cases of intrasubjective, intersubjective and interspecific perceptual variation reveal that the SSR of an object does not fix its perceived color. In most situations, most perceivers of the same species will have approximately intersubjectively uniform color experiences. However, there are always subtle differences in perceived color, and these differences grow less subtle as perceivers cross species boundaries or enter unusual viewing conditions. Sometimes—as with the original image of the dress in Figure 6, which led to newsworthy amounts of bickering on social media worldwide—the differences are stark enough to be practically significant. But the argument for

dissociating colors and SSRs need not rely on dramatic cases like the dress (which can be facilely dismissed as illusory). Even with respect to subtle—but ubiquitous—individual differences in color perception, there is no principled way to settle on one perceived color as *the* veridical representation of the SSR of the object in question, much less as *the* true color of the object itself.

Philosophers have responded to the double dissociation between perceivable colors and the physical properties of colored objects in several ways. Physicalists bite the bullet and claim that one of the perceived colors (or some other color) is the true color of the SSR at hand, and that all others are illusory. So, for example, they might claim that that only the rightmost ball in Figure 3 is really bright orange, that the dress in Figure 6 is really blue and black (even for bees), and that the crocus petals in Figure 8 are really devoid of yellow and green. Idealists about color bite a very different sort of bullet and claim that there is no such thing as veridical color perception, because colors are perceptual figments (Jackson 1977). Finally, irrealists about color bite perhaps the biggest bullet of all and claim that there is no such property as color (Hardin 1988; Boghossian & Velleman 1989).[103]

2.2 Color relationalism

Other philosophers have argued that we need not bite any of these unpalatable

[103] This paragraph does not provide a complete taxonomy of non-relationalist metaphysical theories of color. One notable exclusion is primitivism (Yablo 1995; Gert 2008).

bullets. Relationalists (Thompson 1995; Hatfield 2009; Cohen 2010; Chirimuuta 2015) hold that colors are neither arbitrary intrinsic dispositional properties of objects, arbitrary intrinsic physical properties of objects, purely subjective sense data, nor illusions. Instead, relationalists respond to the double dissociation between perceived colors and SSRs by arguing that colors are constituted by the relationship between color perceivers and the dispositional properties of environmentally embedded objects. Evan Thompson, for example, writes that "being colored a particular determinate color or shade is equivalent to having a particular spectral reflectance, illuminance, or emittance that looks that color to a particular perceiver in specific viewing conditions" (1995: 245). Thompson thus responds to the problem with identifying colors with SSRs by selectively grounding colors in SSRs that bear the right sort of relationship to color perceivers. In particular, colors are realized by those SSRs that enable perceivers to have color experiences. On this relationalist picture, single SSRs sometimes non-arbitrarily realize different colors *because* they enable different perceivers to see different colors (or single perceivers to see different colors in different contexts). Likewise, two different SSRs sometimes non-arbitrarily realize a single color because they both enable perceivers to see that color.

Gary Hatfield articulates a version of relationalism that stresses the biological role of color.

Not every property is a physical property. The property of being nutritious is

> not. Neither is color. They are both biofunctional properties. Color, as a property defined in relation to phenomenal experience or psychological discriminatory capacities, is a psychobiological property. As such, its basis may be found in the relation of subjects to objects. (2009: 296)

To be nutritious is to be usable in metabolism. The physical and chemical properties of any given object, taken by themselves, do not render it intrinsically usable in metabolism. Nothing is non-relationally nutritious; things are only nutritious *for* organisms. Of course, being nutritious for any particular organism is nothing more than a matter of having *appropriate* physical and chemical properties. But what makes those physical and chemical properties appropriate is their metabolizability by that organism. Wood is nutritious for termites but not for humans, and peaches are nutritious for humans but not for termites. The same point applies to color. To be colored is to be perceivable as colored. The physical and chemical properties of any given object, taken by themselves, do not render it perceivable as colored. Things are only colored *for* organisms. Again, being colored any particular hue for any particular organism is nothing more than a matter of having *appropriate* physical and chemical properties. But what makes those physical and chemical properties appropriate is the fact that they allow that organism to perceive that color. Oranges may be orange for (the vast majority of) humans, but that is no guarantee that they are orange for other trichromats, much less for dichromats.

Hatfield calls nutritiousness and color "biofunctional properties". Biofunctional properties exist only in relation to organisms that have evolved to interact with objects by way of those properties. Put differently: biofunctional properties are quintessentially *ecological* properties. They are properties that exist only in the environments of organisms. There is no nutritious wood except in the environments of creatures like termites, and there are no colored objects except in the environments of color perceivers. Biofunctional properties can thus be said to have evolved alongside the organisms in relation to whom they exist. Humans likely evolved trichromatic color vision in part to distinguish fruits from leaves (Mollon 1989; Jacobs 1996; Regan et al. 2001). On this hypothesis, the perceivable colors RASPBERRY RED and LEAF GREEN evolved in tandem with the perceptual capacity of humans to discriminate between how the surfaces of raspberries and leaves respectively reflect light.

Color differs from nutritiousness in that its effect on organisms has a subjective mental component. Hatfield therefore terms color a "psychobiological property". Psychobiological properties are ecological properties that interact with the minds of the organisms whose environments they occupy. Color, realized by SSRs in relation to perceptual systems, is the psychobiological property that functions to enable organisms to visually discriminate environmental objects by hue.[104]

---

[104] The relationalist metaphysics of color I prefer diverges from Hatfield's in one key respect. Whereas I follow Jonathan Cohen (2010) in taking colors to exist relative to particular perceivers and particular viewing conditions—such that the dress, as pictured in Figure 6, really is white and gold for me in that context—Hatfield takes colors to exist relative to standard perceivers (of a given species) and standard viewing conditions (for that species). As Hatfield (personal

It is worth noting that while color relationalists tend to advocate relationalism as an alternative to physicalism about color, it would be perfectly coherent for them to embrace both positions. To see why, recall the distinction between attitudes of belief and cognitive states of belief. Ryleans can find reason to reject the paramechanical conflation of attitudes and cognitive states without thereby endorsing eliminativism about cognitive states. An analogous distinction can be drawn between the colors organisms perceive and the intrinsic properties of objects that physicists might reasonably label 'colors'. Relationalism might be the correct view of color qua perceivable property of objects, even as physicalism is true of color qua intrinsic property of objects. (Perhaps relationalism is the correct account of the 'color' phenomenon of interest in perceptual psychology, whereas physicalism is the correct account of the 'color' phenomenon of interest in the physics of light.)

Relationalists need not take a stance on the issue of whether SSRs resemble (or correspond to) perceivable colors sufficiently to count as the intrinsic colors of physical objects. Instead, they can point to the double dissociation between SSRs and perceivable

---

communication) points out, Cohen's relativistic view of color has the unintuitive consequence that everything is black in the dark. Cohen (2010: Chapter 4) responds to this objection by developing a semantics for color attribution that diverges from his relativistic metaphysics of color. In my estimation, Hatfield's view runs into greater difficulties trying to provide principled grounds for determining what renders perceivers and viewing conditions 'standard', and, more importantly, trying to provide principled grounds for determining what makes being standard metaphysically operative in constituting color. However, I will not flesh out those objections here. My relativism about attitudes of belief—while usefully illustrated by analogy with Cohen-esque relativism about color—does not stand or fall with any particular metaphysics of color. Hatfield may well have the best view of color, even if my arguments in Chapter 7 successfully defend relativism about attitudes of belief.

colors as reason to reject physicalism and embrace relationalism as an account of *perceivable* color in particular. I endorse the relationalist view that perceivable colors are psychobiological properties relativized to particular perceivers and particular viewing conditions, while remaining agnostic about whether SSRs or their physico-chemical realizers deserve to be labeled 'intrinsic colors' in their own right.

The rest of this chapter will exploit an analogy between belief and color in order to resist the conflation of attitudes of belief and cognitive states of belief, and argue for minimal Ryleanism about the former. According to Ryleans, attitudes of belief are psychobiological properties, just like perceivable colors. Just as perceivable colors are realized by SSRs that function to make objects look colored to color perceivers, attitudes of belief are realized by dispositions to act, react, think and feel that, taken together, function to render people *believers* in the eyes of belief attributors.[105]

---

[105] Like all analogies, the analogy between belief and color is imperfect and incomplete. Most strikingly, we directly see colors. Following Shannon Spaulding (2015), I reject views of mindreading according to which we literally perceive beliefs. In some ways, this point of disanalogy renders the metaphysics of attitudes of belief easier to pin down than the metaphysics of perceivable colors. Embracing belief attribution as a thoroughly cognitive (as opposed to perceptual) phenomenon enables us to ignore tricky questions about the cognitive penetration of perception. It also enables model-theoretic accounts of belief attribution (see Chapter 3.3); cognitive models of belief are easier to get theoretical and empirical grasps on than the amodal perceptual processes posited by theorists who countenance mindseeing. A related point of disanalogy stems from the impact that cultural forces can have on models of belief (see Chapter 7). While culture may affect color perceptual learning (Connolly 2017), cognitive models of belief are likely much more culturally variable (see Chapter 7.3 for discussion of the empirical literature concerning cultural effects on models of belief). Likewise, belief attributors can construct models of new beliefs on the fly, whereas color perceivers cannot learn to perceive new colors.
The final point of disanalogy I will mention is that perceivable colors are discrete properties—fully explicable without reference to any other properties (e.g. size) of colored objects. Attitudes of belief, on the other hand, are aspects of molar personalities—fully explicable only with reference to other characteristics of whole believers (see Chapter 6.3.3).

3. Belief

There is a problem for paramechanical views that conflate attitudes of belief with cognitive states of belief that parallels the problem for physicalism about perceivable color. The primary candidates for cognitive states of belief are brain states, subpersonal computational states, and personal purely functional states. But there are double dissociations between attitudes of belief and all three of these candidates for cognitive states of belief, which mirror the double dissociation between perceivable colors and SSRs.[106]

3.1. Against conflating attitudes with brain states

Paramechanical mind-brain identity theorists identify attitudes of beliefs with brain states or processes; as JJC Smart puts it, beliefs "just *are* brain processes, not merely *correlated with* brain processes" (Smart 2007: 1).[107] Unfortunately for paramechanical identity theorists, there is a much-rehearsed double dissociation between attitudes of beliefs and the relevant neural processes and states of believers.

First, distinct types of brain state correspond with identical types of attitude of

---

[106] The dissociations I present in this section have a cascading structure: although paramechanical computationalisms dodge the arguments against paramechanical mind-brain identity theories, and paramechanical pure functionalisms dodge the arguments against paramechanical computationalisms, the arguments against paramechanical pure functionalisms also condemn paramechanical computationalisms, and the arguments against paramechanical computationalisms also condemn paramechanical mind-brain identity theories.

[107] The classic mind-brain identity theorists focused on identifying states of consciousness, rather than attitudes, with brain states. Indeed, U.T. Place (1956) developed his pioneering version of the identity theory about consciousness as a supplement to his staunch Ryleanism about belief.

belief. Even granting the identity theorist's thesis that token beliefs are identical to token brain states, there is no reason to assume that the type-individuated attitude of belief that water is potable (for example) is neurally implemented in the exact same way in different believers. Even philosophers who are skeptical of the prevalence of multiple realiz*ation* tend to admit that beliefs are, in principle, multiply realiz*able* (Shapiro 2008: 514). And mere multiple realizability (even in the absence of actual multiple realization) is problem enough for paramechanical versions of the identity theory. Even if, as a matter of contingent fact, attitudes of belief perfectly track human brain states, there would be no principled reason to claim that any particular brain state is the single proper brain state for the belief it happens to realize. (There would only be reason to claim that human nervous systems just so happen to realize that belief with that brain state.) Moreover, attitudes of belief are very likely multiply realized across species boundaries; my belief that water is potable is likely realized by a very different brain state than my neighbor's chicken's belief that water is potable, given (among other things) that my brain includes a neocortex and the bird's brain does not (cf. Karten 2013; Polger & Shapiro 2016: 114–117).

Second, on the least controversial of externalist assumptions, indistinguishable brain states can underlie distinct attitudes of beliefs when embedded in distinct environments. Externalists take attitudes of beliefs to be individuated, in part, with respect to the external environments of believers. Thus, Putnam's (1975) Earthly believer believes that *water* (H2O) is potable, whereas his exact physical duplicate on Twin Earth

believes that *t-water* (XYZ) is potable. The Earthly and Twin Earthly believers are physically identical yet have different attitudes of belief. Brain states thus do not fix attitudes of belief. There is no principled way of picking one attitude of belief—say the belief that water (rather than t-water) is potable—as *the* proper belief for the brain state underlying both beliefs.

Perhaps *cognitive states* of belief just are brain states. But the double dissociation arising from multiple realizability and externalism indicates that *attitudes* of belief should not be conflated with brain states. Insofar as the identity theory is the best metaphysics of cognitive states of belief—and a recent philosophical revival provides good reason to think it may be (Shapiro 2018; Thomson & Piccinini 2018)—we ought not conflate attitudes with cognitive states.

Paramechanical philosophers have responded to this double dissociation between attitudes of belief and brain states in several ways. Persistently paramechanical mind-brain identity theorists downplay multiple realization and either deny externalism or broaden the relevant physical states to include features of the environment. Dualists pull apart the physical and mental domains. Eliminativists deny that beliefs exist. However, by far the most popular way to respond to the double dissociation is to adopt paramechanical functionalism about belief. Paramechanical functionalism encompasses a diverse family of views that all hold that beliefs are individuated with respect to the functional roles they play in cognitive systems.

3.2. Against conflating attitudes with subpersonal computational states

Most functionalists can be (roughly) categorized as computational functionalists.[108] Cognitive systems receive inputs and emit outputs. Computational functionalism is the doctrine that beliefs (alongside desires and, on some accounts, intentions) are subpersonal cogs in cognitive machines that play the functional role of transforming perceptual inputs into behavioral outputs. In other words, beliefs are nothing more and nothing less than functionally characterized computational states.[109] Realizer functionalists identify beliefs with the brain states that realize these computational roles. Role functionalists identify beliefs with the computational roles themselves. According to the currently most popular version of role computational functionalism—subpersonal psychofunctional representationalism—beliefs are assertive psychofunctional relations to mental representations (Fodor 1987; Quilty-Dunn & Mandelbaum 2017).

All computational functionalists embrace the multiple realizability of computational roles by brain states. For the functionalist, brain states are able to

---

[108] I'm using the term 'computational' loosely. 'Computational functionalisms', as I use the term, include machine state functionalism, most classic forms of psychofunctionalism, and many embodied/embedded/extended/enactive/etc. functionalisms that retain some commitment to mental states intervening between perception and action. Computational functionalism is, however, to be differentiated from pure functionalism, which I discuss in Section 3.3., and teleofunctionalism, which I discuss in Section 4.

[109] Note that the subpersonal computational states discussed here and in Section 3.2. exist at Marr's (1982) 'algorithmic' rather than 'computational' level (as explained in Chapter 2). What Marr calls the 'computational' level instead plays host to the personal pure functional states I will discuss in Section 3.3.

multiply realize attitudes of beliefs because what makes a brain state realize a belief is a matter of cognitive function rather than precise physical characteristics. Many computational functionalists are also happy to extend cognitive systems into the external environment, and therefore embrace the externalist idea that environmental factors can help individuate functional roles (Harman 1987; Block 1990; Kitcher 1991). Many varieties of computational functionalism thus elegantly dodge the double dissociation between brain states and attitudes of belief while reaffirming the paramechanical conflation of attitudes of belief with cognitive states of belief. Nevertheless, paramechanical computational functionalists must reckon with a double dissociation of their own.

First, attitudes of belief are multiply realized by subpersonal computational states, just as computational states are multiply realized by brain states. Consider three people who all believe that a mug contains hot coffee. Delia orders a "dark roast for here" from the barista at her local café, and tells her three-year-old son, Roger, "watch out, Mommy's coffee is very hot." Pouring the coffee, the barista mentally represents a 196°F French Roast made from Guatemalan beans being poured into a stoneware mug made by a local potter. Delia's relevant mental representations are comparatively impoverished; to her, it is a dark roast coffee that she remembers liking the taste of, being poured into a nice mug, which is steaming enough that she should warn Roger that it is very hot. For his part, Roger, who has been raised alongside his mom's coffee habit, takes 'mug' to refer to any drinking vessel whatsoever, has no clue that there are

different kinds of coffee, and has only a rough conception of different degrees of heat.

The barista, Delia, and Roger all share the attitude of belief that there is hot coffee in the mug. But the functionally individuated computational states that underlie their respective beliefs diverge sharply, in terms of both representational content and psychofunctional role. Roger's notions of 'mug' and 'coffee' and 'hot' are different than Delia's, though not so different that it results in them having a different belief in this context.[110] Moreover, Roger's subpersonal computations over representations of the coffee and mug transform different inputs into different outputs than his mom's relatively richer representations, and the barista's still richer representations transform still different inputs into still different outputs. For example, the barista alone is prone to feel proud that the coffee is the optimal temperature for this particular Guatemalan French Roast, and Roger alone is prone to infer that every mug, cup, and glass in the café is a mug containing scary hot coffee. The three café denizens' subpersonal computational states multiply realize their shared attitude of belief.[111]

The case for the multiple realization of attitudes of belief by subpersonal

---

[110] A thoroughgoing externalist (Burge 2010 or Fodor & Pylyshyn 2016, for instance) might deny this difference in representational content. However, in my view, which I cannot argue for here, all plausible externalisms allow for internal factors that cause differences in content in mental representations with the same referent. Regardless, the claim that Delia, Roger, and the barista's mental representations have different content is not required to establish the dissociation between attitudes of belief and computational states: the differences in the psychofunctional roles—or what Fodor calls "the syntactic structure of modes of presentation" (1992: 54)—are enough to do the trick.

[111] Ryle (1949), Dennett (1978), Pylyshyn (1980; 1984), and Schwitzgebel (2018) use similar cases to plump for the same conclusion: subpersonal computational states multiply realize personal attitudes.

computational states is strengthened by cross-species comparisons. As Kim Sterelny (1990a, 1990b) argues, animals of different species often subpersonally realize the same attitudes in dramatically distinct computational states.

> Consider just the difference in perceptual structures between bats and owls. Owls have notoriously acute night vision, whereas bats find their way around by echo location. So if we had reason to attribute to bats and owls the same psychological state—say that they both perceive mice—then that state could hardly be individualistically defined. Perceptual systems vary greatly; their only common feature is that their function is the extraction of information for the adaptive control of behavior. (Sterelny 1990b: 98)

As with the state of perceiving mice, a bat and an owl can share the attitude of believing there is a mouse in the grass, despite great variance in the computational states underlying the two creature's respective beliefs. There is no principled way of picking the bat or owl or, for that matter, forest ranger's subpersonal relations to relevant mental representations as *the* proper computational state to realize the belief that there is a mouse in the grass.

Second, attitudes of belief exhibit a unity that may not exist among the computational states that realize them at the subpersonal level. Consider my attitude of belief that I can help myself to free La Colombe coffee from the Center for Neuroscience

and Society lounge on weekday mornings. This belief is concrete, specific, and coherent, but its cognitive underpinnings are complex. Believing it requires a cognitive system featuring a conjunct of functionally independent computational relations to mental representations, including representations of coffee and the company La Colombe and freeness of charge and permission and weekday mornings and my capacity to fetch things, as well a mental map of how to get to the CNS lounge.

There is unity in my attitude of belief—I live as if I can help myself to free La Colombe from CNS on weekday mornings in a unified pattern—but this unity plausibly emerges from disunity at the subpersonal level of computational states. My psychofunctional relations to my mental representation of weekday mornings is only peripherally associated with my psychofunctional relations to my mental representation of freeness of charge. It is dubious that an explanatorily fecund account of cognitive architecture would non-arbitrarily tie these functionally independent computational states together as components of a single, functionally unified subpersonal cognitive state of belief. The only reason to tie them together would be that they both underlie an attitude of belief. But that belief should not be conflated with a conjunction of computational states that are disunified at the subpersonal level; the whole emergent pattern is more than the sum of its computational realizers.[112]

Especially in light of the reasons provided in Chapter 3 to doubt that the folk construe beliefs as inner causes, the multiple realization of attitudes of belief by

---

[112] Matthews (2007: 241) makes much the same point.

(plausibly functionally disunified conjunctions of) subpersonal states should be enough to convince theorists to refrain from conflating attitudes with their computational realizers. And it has so convinced some paramechanists. Sterelny, for example, avoids the double dissociation between attitudes of belief and subpersonal computational states by instead conflating attitudes of belief with cognitive states of belief that play functional roles at the personal level of explanation. For high-level paramechanical functionalists like Sterelny, the bat, owl, and forest ranger all believe that there is a mouse in the grass insofar as they are disposed to act, react, think, and feel in patterns that function to track the existence of the mouse in the grass. Paramechanists like Sterelny require that "the animals in question have certain discriminatory, memory or calculative abilities [to qualify as believers], but don't care how those abilities are computationally realized" (1990b: 99).

3.3. Against conflating attitudes with personal purely functional states

Frank Jackson and Philip Pettit call this a "purely functional theory", since it dictates that "to have beliefs and desires is to be understood purely functionally" (1990: 43): as having no relevant features over and above (or under and below) delivering certain outputs in response to certain inputs. Like Sterelny, Jackson and Pettit stress that the relevant functional roles mediating between inputs and outputs are *not* subpersonal psychofunctional roles of the sort that differentiate owls, bats, and forest rangers (34, 37). Instead, they are personal patterns of dispositions described at an abstract enough level

to subsume the ways in which owls, bats, and forest rangers function to represent their environments. It does not matter whether disunified conjuncts of computational states multiply realize these abstractly characterized and purely functionally individuated beliefs; "it does not matter for the success of our passage back and forth between situations, behavior, beliefs, and desires how many states inside the agent are required to work the trick" (41). Insofar as agents do successfully work the trick—do go from inputs to outputs in patterns reliably identifiable as believing—those agents are believers.

On this (purportedly) purely functional person-level conception of cognitive states of belief, Delia, Roger, and the barista all share the cognitive state of belief that there is hot coffee in the mug, and there is functional unity to my cognitive state of belief that I can help myself to free La Colombe coffee from the Center for Neuroscience and Society lounge on weekday mornings. Moreover, as Jackson and Pettit argue, it is extremely difficult to deny that such personal purely functionally individuated beliefs exist.

The problem is that there is no such thing as a *purely* functional analysis of agents, devoid of a principled criterion which determines which mediations between inputs and outputs are functionally relevant. In other words, the person-level functionalist must provide a criterion that fixes which of a believers' countless dispositions make up any given belief; something has to explain why Delia's disposition to blow softly into her mug—but not her disposition to sneeze when placing her mug

next to a vase of daisies—is partly constitutive of Delia functioning to represent the mug as containing hot coffee.

The diverse styles in which believers can play person-level functional roles put this problem in high relief. The echo-locating bat and sharp-eyed owl are disposed to transform different outputs into different inputs, yet their divergent dispositions functionally realize the same belief. There are stark intraspecific differences in style of belief as well. Delia, Roger, and the barista possess different personal dispositions as well as different subpersonal computational states. Their divergent dispositions functionally realize the same belief, but they do not *purely* functionally realize the same belief. The aspiring person-level functionalist must give a criterion that non-arbitrarily lumps Delia, Roger, and the barista's distinctly stylized patterns of living together as fulfilling the same function for the respective believers.

Any of three candidate criteria would serve to explain what makes divergent patterns of dispositions all fulfill a particular person-level functional role, and thereby count as the same belief. First, divergent patterns of dispositions might all stem from the same subpersonal computational state (or set of computational states). In that case, person-level paramechanical functionalism collapses into subpersonal paramechanical functionalism and faces the attendant double dissociation between attitudes of belief and cognitive states of belief. Second, divergent patterns of dispositions might all serve the same irreducibly normative ecological purpose. Sterelny opts for this second, teleofunctional criterion. I will discuss teleofunctionalism—and argue that it should be

wielded in support of the distinction between attitudes of belief and cognitive states of belief—in Section 4. Third, divergent patterns of dispositions might all sufficiently fit folk models of belief. Jackson and Pettit opt for a paramechanical version of this third, commonsense criterion, arguing that "it is sufficient for having beliefs and desires that one be in states which satisfy the functional roles embodied in our everyday practice of predicting and explaining human behavior (for short, the folk roles)" (1990: 36).[113]

Despite appearances, this commitment to models of beliefs providing the criteria that fix beliefs does not make Jackson and Pettit Ryleans. On the contrary, Jackson and Pettit hold that "folk psychology is a *theory*" about purely functional person-level cognitive states that exist independently of models of belief (33). People have myriad cognitive states that nomically mediate between inputs and outputs. On Jackson and Pettit's view, folk models of belief simply determine *which* independently existing cognitive states—which lawlike mediations between inputs and outputs—are captured by the commonsense theoretical positing of "beliefs". Jackson and Pettit embrace the paramechanical conclusion that attitudes of belief are just person-level, purely

---

[113] Dan Singer (personal communication) has suggested (without endorsing) a fourth option: it might just be a brute fact that certain functionally individuated patterns of dispositions constitute certain cognitive states of belief. Paul Churchland has rightly dismissed this "functionalist stratagem" as "reactionary, obfuscatory, reactionary, and wrong" (1981: 81); it would prove too much, as parity of reasoning would entail realism about things like phlogiston qua 'whatever plays a functional role in combustion'. Jackson and Pettit concur (1990: 51). In addition, the paramechanical strategy of identifying beliefs with whatever ways people turn out to functionally represent their environments would fall prey to the same dissociation between attitudes of belief and cognitive states of belief that I marshal against Jackson and Pettit's folk role criterion.

functionally individuated cognitive states that folk belief attributors designate "beliefs".

Unfortunately for paramechanical commonsense functionalists like Jackson and Pettit, there is a final double dissociation between attitudes of belief and the person-level functional states picked out by models of belief. Just as SSRs—surface-level functional states of objects—multiply realize perceivable colors in cases of metamerism, person-level pure functional states multiply realize attitudes of belief. And just as the perceived colors of objects (but not SSRs) vary between perceivers, attitudes of belief (but not person-level pure functional states) vary between belief attributors.

First, Jackson and Pettit's commonsense paramechanical functionalism fails to make the case that Delia, Roger and the barista all share a cognitive state of belief. The three café denizens all have the same attitude of belief, as they are all in states which satisfy the folk role played by the attitude of believing there is coffee in the cup. According to Jackson and Pettitian commonsense functionalism about cognitive states of belief, the three also all have cognitive states of belief: whatever person-level dispositions happen to mediate between inputs and outputs such that they function in the manner theoretically posited by the folk role. But Delia, Roger, and the barista do not share the same cognitive state of belief; on the contrary, they share an attitude of belief *despite* having different stylized person-level functional configurations, just like metamers share a color despite boasting different SSRs.[114]

---

[114] I elaborate how Ryleans—and Ryleans alone—can adequately account for styles of attitude of belief—such that two people can both have the same attitude of belief and believe it in dramatically different styles—in Chapter 6.

Attitudes of belief are individuated based on belief attributors' models of belief, whereas cognitive states of belief are individuated within the functional structure of believers. Jackson and Pettit (1990: 40–43) make a plausible case that the predictive and explanatory success of attributions of attitudes of belief guarantees that people have cognitive states of belief—that ascriptions of attitudes of belief track some way or other in which believers are functionally configured. But there is no guarantee that all ascriptions of any given attitude track the same functional configuration. On the contrary, personal pure functional states multiply realize attitudes of belief. There is no principled way of picking Delia, Roger, or the barista's purely functional state as *the* proper person-level functional state that is veridically captured by an ascription of attitude of belief.

Nor does the purely functional configuration of a believer fix what attitudes of belief she possesses. In some contexts, it is intersubjectively indeterminate what somebody believes, just as the colors of the dress in Figure 6 and crocus in Figure 8 are intersubjectively indeterminate. A single personal functional configuration sometimes underlies different attitudes of belief for different belief attributors. In these cases, the believer believes one thing for one belief attributor and another thing for another belief attributor, just as the dress is gold and white for one color perceiver and black and blue for another color perceiver. There is no principled way of settling on one of these

attitudes as the believer's one true belief.[115]

Recall, from Chapter 1, Dennett's (1998: 115) vignette about Ella, who has been behaving in troubling, self-undermining ways. Brown interprets Ella as believing that she ought to kill herself; Jones interprets Ella as believing, despite her angst, that she ought not kill herself. Both Brown and Jones, by stipulation, have access to the full range of Ella's dispositions to act, react, think, and feel. Even so, they disagree. It is intersubjectively indeterminate what Ella believes.

To flesh out the case, suppose Ella's most relevant dispositions are threefold: she feels no joy, it regularly occurs to her that she is capable of killing herself, she says she wants to die. Now suppose that Brown and Jones have developed different interpretive strategies over the course of their lives, which lead them to construe their models of belief differently. Through his amateur study of human psychology and 19th century debates about natural selection, Jones has developed the deep conviction that people never believe that they should kill themselves. Jones is convinced that evolutionary pressures have rendered people psychologically incapable of believing that they should

[115] It is worth noting at the outset of this discussion that Ryleans need not accept the relativistic view I pursue; this relativism is not entailed by the commitments of minimal Ryleanism. Indeed, no other Rylean does accept it, though I will argue that they ought to in Chapter 7. Nevertheless, the possibility of strong intersubjective indeterminacy—embraced by my relativistic Ryleanism—yields the strongest dissociation between attitudes of belief and all candidates for cognitive states of belief, and thus provides Ryleans with their strongest argument against paramechanists. Moreover, as argued in Section 4, the analogy with perceivable color motivates relativism; how beliefs arise in the environments of belief attributors inevitably varies along with the idiosyncrasies of the models wielded by those belief attributors, just as perceivable colors vary along with the idiosyncrasies of color perceivers' perceptual systems.

end their own lives, though he allows that people sometimes mistake other beliefs—such as the belief that they would do anything to escape their depression—for the belief that they ought to kill themselves. Jones models Ella as being confused about what she believes (and worries that she might act on the basis of that confusion), but does not model Ella as believing that she ought to kill herself.[116] Brown, in contrast, interprets people as believing the things they assert unless he has probative evidence to the contrary, and takes it for granted that people often have genuine suicidal beliefs.

Jones voluminously and reliably predicts behaviors, thoughts, and feelings in accordance with his pet psychological theory (just look at how many people unsuccessfully attempt suicide, due to the lack of conviction with which they act!). Brown voluminously and reliably predicts behaviors, thoughts, and feelings in accordance with his no-bullshit ethos. Ella believes she ought not kill herself from Jones's point of view, while believing that she ought to kill herself from Brown's point of view. Jones and Brown attribute different attitudes of belief despite agreeing on Ella's

---

[116] You might think that Brown's is a deeply flawed way of thinking about people's motivations. By stipulation, it is a predictively powerful interpretive strategy. (Brown predicts whether or not people will kill themselves with as much accuracy as Jones; he just doesn't use the attribution of belief that one ought to kill themselves in order to get to that prediction.) But you might think that it fails to capture what people *actually* believe; what actually motivates them to act. However, whether Brown is a non-ideal social cognizer is beside the point. Insofar as attitudes of belief are determined by folk models of belief—as Jackson and Pettit readily admit they are—they are determined by the non-ideal, messy ways in which belief attributors actually model beliefs. Or so I will argue in Chapter 7. For now, compare: you might think that humans are deeply flawed color perceivers. We fail to represent whole chunks of the spectrum! But that humans are non-ideal color perceivers is beside the point, when it comes to the metaphysics of perceivable colors. Perceivable colors are determined by the non-ideal, messy visual spectrum, as it manifests itself in relation to particular non-ideal color perceivers.

purely functional person-level states.

What does Ella believe, *really*? Well, what color is the dress, *really*? (If your kneejerk reaction is "it is really blue and black, just look at Figure 7!", then what precise shades of blue and black is it? And is it really blue and black for bees as well as humans?) These questions are insufficiently precise. Ella believes different things—and believes things differently—for different belief attributors. The dress is blue and black for 30% of the population and white and gold for 60% of the population. (If you still think Figure 6 is a special case: the dress is a different shade of blue for different color perceivers. And crocuses are different colors for different animals.) Everybody is right. The dress *really* is white and gold for me, and it *really* is blue and black for people who see it as blue and black.[117] Likewise, Ella *really* believes he ought not kill herself for Jones, while *really* believing she ought to kill herself for Brown.

Knowledge of all of Ella's dispositions does not suffice to decide between these two interpretations, any more than knowledge of the dress's SSR suffices to decide its precise shade of perceivable blue. Ella has a single, intersubjectively stable purely

---

[117] As suggested in the parentheticals in this paragraph, readers need not accept this individualist relativism about perceivable colors in order to accept my analogous individualistic relativism about attitudes of belief. It may be that the dress is *really* particular determinate shades of blue and black for all humans—determined by species-standard perceivers and viewing conditions (see fn. 104, above)—even while attitudes of belief are fully intersubjectively indeterminate. The analogy between belief and color is meant to illuminate the view of belief I have advanced, but it does not stand or fall with any particular relationalist metaphysics of color. In Chapter 7, I will argue against the ways in which other Ryleans—like color relationalists who relativize colors to standard perceivers and viewing conditions rather than particular perceptual systems—have set various normative standards to intersubjectively determine attitudes of belief.

functional set of dispositions to act, react, think, and feel, but she has different attitudes of belief for Brown and Jones.[118] There is no principled way to settle on one of these attitudes of belief as *the* veridical representation of Ella's purely functional cognitive state, much less as Ella's one true belief.

As with color perception, there is interspecific as well as interpersonal indeterminacy in attitude attribution. Objects do not have the same perceivable colors for dichromats as for trichromats. Analogously, mindreadees do not have the same attitudes for nonhuman mindreaders as for humans. According to the currently mainstream interpretation of the empirical evidence on chimpanzee mindreading capacities, chimps attribute goals, intentions, perception, knowledge, and ignorance to other animals, but do not attribute attitudes of belief (Call and Tomasello 2008). Even if recent reports that chimps do attribute attitudes of belief have merit (Krupenye, Kano et al. 2016), it is unlikely that they model beliefs in the same manner as humans. For example, chimps almost certainly do not model beliefs as centrally involving the

[118] It might be thought that knowledge of Ella's dispositions is sufficient for deciding what she in-between believes (Schwitzgebel 2001), and that cases of intersubjective indeterminacy are really just cases of in-between believing (in which two interpreters attribute different beliefs that the in-between believer is in-between believing). Perhaps Ella only kinda, sorta believes that she should kill herself. However, this reading changes the case. Intersubjective indeterminacy cannot be chalked up to the misattribution of full belief in cases of in-between belief. Indeterminacy arises due to competing models of belief, whereas in-between believing arises when a believer possesses some subset of the dispositions a particular model construes as realizing belief. If Ella objectively in-between believed that she ought to kill herself, then both Jones and Brown (given their stipulated local omniscience) would know that Ella had some but not all of the dispositions that make up the belief that she ought to kill herself and would accordingly agree that she in-between believes. Instead, Jones and Brown disagree about which dispositions make up which beliefs and are thus faced with true intersubjective indeterminacy.

disposition to linguistically assent to certain propositions.

This difference in models of belief does not entail that chimps always fall short of veridically attributing the beliefs that believers *really* have. Instead, chimps (like humans) may be excellent at attributing the attitudes it serves their species-specific purposes to attribute (Boesch & Boesch-Achermann 2000; Boesch 2009; Lurz 2009; Nagel 2011; Andrews 2012; Andrews 2018). There is no interspecifically determinate answer to the question 'what attitude does Chimpella *really* have?', just as there is no interspecifically determinate answer to the question 'what perceivable color does the crocus *really* have?' Chimpella really has one attitude of belief for Sally the human, and a different attitude for Red Peter the chimp. Of course, Chimpella does not have an interspecifically variable person-level purely functional state of belief. Thus, Chimpella's relevant attitude(s) ought not be conflated with her relevant cognitive states of belief.[119]

The reader need not accept my full-on relativism about attitudes of belief in order to see the problem intersubjective indeterminacy presents for conflating attitudes of belief with cognitive states of belief. An anti-relativist could, instead, treat intersubjectivity as grounds for skepticism rather than relativistic realism about attitudes of belief.[120] Or else the anti-relativist might follow Dennett (1998) and claim that what Ella believes is *objectively* (rather than merely *intersubjectively*) indeterminate. Finally, the anti-relativist might follow Davidson (2001), Baker (1995), and Mölder (2010) and deny

---

[119] I discuss the comparative psychology literature on belief attribution at length in Chapter 8.
[120] Figure 2 made the distinction between paramechanical eliminativism and attitude skepticism (which refrains from conflating attitudes with cognitive states while denying that attitudes exist).

the coherence of the Ella and Chimpella cases—indeed, the coherence of intersubjective and interspecific indeterminacy—on the grounds that veridical belief attribution is an irreducibly social, irreducibly normative, and uniquely human ability.[121] All of these theorists can agree that Jones and Brown (or Sally and Red Peter) attribute different attitudes of belief to (Chimp)Ella, and that there are no facts to be uncovered *about (Chimp)Ella's functional architecture* that fixes one attribution as *the* veridical attribution.[122] Attitudes—if they exist—are metaphysically fixed by interpretive capacities or practices rather than pure person-level functional roles, subpersonal computational states, or brain states. Moreover, brain states multiply realize computational states which multiply realize pure functional roles which multiply realize attitudes of belief. Thus, attitudes of belief ought not be conflated with cognitive states of belief.

3.4. How attitudes of belief are like perceivable colors

My Rylean solution to the problems facing paramechanical theories of belief mimics the relationalist solution to the problem facing physicalist theories of color.

---

[121] For Davidson, it follows from the irreducibly social nature of belief attribution that if Jones and Brown know the same set of facts about Ella, then they cannot rationally disagree about what Ella believes. For Baker and Mölder, Ella believes whatever the community standard interpretation would have her believe. I will argue for relativism, against these other versions of Ryleanism, in Chapter 7. The present point is that attitudes ought not be conflated with cognitive states if attitudes are determined relative to models of belief—whether the veridicality conditions for belief attribution are set relativistically, objectively, or intersubjectively.

[122] For attitude skeptics, this is because no lay belief attributions are veridical. For Dennettians, this is because there is no objective fact of the matter. For Davidsonians, this is because what fixes the veridical attribution is a constitutive norm of interpretation—the principle of charity—not a cognitive state of belief. For Bakerites and Mölderians, this is because what fixes the veridical attribution is the community standard attribution, not a cognitive state of belief.

Attitudes of belief, like perceivable colors, are relational properties. Just as perceivable colors arise in the relation between colored objects and color perceivers, attitudes of belief arise in the relation between believers and belief attributors. In particular, believing is possessing a pattern of dispositions that a belief attributor recognizes as a taking of the world to be some way. Importantly, it is the *pattern* of dispositions—which emerges as an attitude of belief only in relation to a belief attributor's model—rather than the purely functionally individuated dispositions themselves, that belief attributors recognize as beliefs.

Thus far, this is just minimal Ryleanism about attitudes of belief: the sparse framework of a view that (as I argued in Chapter 1) is agreed upon by all philosophers articulating alternatives to paramechanical views of belief. But the analogy with color also indicates a novel way of beginning to flesh out this minimal Rylean framework. Attitudes of belief are psychobiological properties, like perceivable colors. Beliefs serve functional roles in the environments of mindreaders and metacognizers. In particular, attitudes of belief are the externalistically individuated properties in virtue of which believers—qua objects in social environments—function to seem to believe to belief attributors—qua subjects making sense of the objects in their social environments. Because attitudes of belief exist only in relation to organisms who have evolved to interact with their social environments by way of the attribution of belief, they are biofunctional properties.[123] Because their effects on attributors have a subjective mental

[123] Much more on this evolutionary story is coming in Section 4.

component, attitudes of belief are psychobiological properties. To believe that a mug contains hot coffee is to function, as an object in a social environment, to be attributed the belief that a mug contains hot coffee. Functioning in this way is a matter of having an appropriate pattern of dispositions to act, react, think, and feel: that is, a pattern of dispositions that the belief attributor recognizes as constitutive of belief.

Paramechanists are mistaken to assume that attitudes of belief function as cogs in the cognitive systems of believers. Attitudes do not contribute directly to the cognitive operations of the believers to whom they are attributed.[124] Cognitive states of belief do that causal work. Attitudes of belief, by contrast, function as characteristics of believers (including the belief attributor herself) in attributor-relative social environments.

Diverse styles of belief and intersubjective indeterminacy pose problems for paramechanists, but not for Ryleans who refuse to conflate attitudes of belief with cognitive states of belief. Despite their disparate cognitive states, Delia, Roger, and the barista all share the attitude of belief that there is coffee in the mug, as they all live in ways that sufficiently fit the general-purpose model of that belief that I wield as a belief attributor.[125] And, as previously mentioned, Ryleans afford themselves the flexibility to

---

[124] Attitudes of belief do, of course, contribute indirectly to cognitive operations. For example, the bolt of self-knowledge that comes along with attributing oneself the attitude of belief that one's job sucks might lead one to quit their job.

[125] The barista—unlike Delia and Roger—also believes that there is coffee in the mug in a particular, knowledgeable style captured by a believer-specific model of her belief. This difference in beliefs seen in a specialized metaphorical light is analogous to the difference in color between the metamers seen in orange light in Figure 4. For much more on the difference, and interplay, between general-purpose and believer-specific models, see Chapters 3 and 6.

account for—or dismiss—intersubjective indeterminacy in any of several ways. Dennett embraces (rare and practically insignificant) objective indeterminacy; Davidson insists that what somebody believes is objectively and intersubjectively determined by the principle of charity; Baker and Mölder advert to the folk psychological consensus. The relativistic Ryleanism I will defend in Part II embraces the intersubjective indeterminacy of attitudes of belief while denying that there is any objective indeterminacy about what somebody believes. Ella objectively believes one thing for Brown and objectively believes another thing for Jones.

Ryleans thus dodge the concerns about multiple realizability, externalism, and intersubjective indeterminacy that plague paramechanical versions of the identity theory, computational functionalism, and pure functionalism. However, I have not yet fully vindicated minimal Ryleanism. Another popular version of the view that attitudes of belief are theoretically posited cogs in cognitive systems—paramechanical teleofunctionalism (Millikan 1984; Lycan 1987; Dretske 1988; Sterelny 1990b; Papineau 1993; Burge 2010; Neander 2017)—dodges these concerns almost as deftly as Ryleanism.

4. Evolution

The central insight of teleofunctionalism is that beliefs and other content-bearing mental states have irreducibly normative functions. According to paramechanical teleofunctionalists, beliefs *evolved* (and, in individual believers, develop) to be cogs in cognitive systems that serve representational and inferential purposes.

4.1. Paramechanical teleofunctionalism about belief

Whereas Ryleanism is the theory that attitudes of belief are patterns of dispositions that belief attributors identify with taking the world to be some way, paramechanical teleofunctionalism is the theory that attitudes of belief are cognitive states that serve to make representations of believers' environments available for cognitive processing. The dominant, etiological form of teleofunctionalism dictates that beliefs have been naturally selected to serve the evolutionary purposes played by believers' cognitive systems.

Teleofunctionalism differs from other versions of functionalism by placing an emphasis on the biological (as opposed to computational or pure dispositional) functions played by beliefs. As Elliot Sober has quipped, teleofunctionalists undertake the task of "putting the function back into functionalism" (1985: 175). Computational roles and dispositions are not usually described in normative terms; instead, the functionalists discussed in Section 3 make purely descriptive claims about how beliefs function to transform inputs into outputs. Teleofunctionalists, on the other hand, make irreducibly normative claims about how beliefs are supposed to function, given their evolutionary (and developmental) etiology.[126] An attitude or cognitive state has a teleofunction insofar as it is constitutively aimed at fulfilling a biological purpose of an

[126] There are difficult questions about the relative roles of phylogeny and ontogeny in the assignment of teleofunctions. (Most teleofunctionalists focus almost entirely on phylogeny, although Dretske, at least in his later work, locates the teleology relevant to the individuation of beliefs entirely in ontogeny.) However, because it is not my intention to argue either for or against teleofunctionalism, I am going to set these difficult questions aside.

organism.[127]

According to Ruth Millikan's influential etiological teleofunctionalism (1984, 1989, 1993), for example, beliefs and other mental attitudes are the evolved cognitive means by which organisms assimilate information from their environments and figure out how to behave accordingly. Millikan writes that "the categories of intentional psychology are function categories in the biologist's sense of 'function', taking this to be a sense in which function is determined by evolutionary history rather than by current dispositions" (1993: 171). What beliefs you have does not depend solely on the current computational (or purely functional) architecture of your cognitive system; it depends on the naturally teleological etiology of the various mechanisms and cognitive states that system comprises.

> This means that categories such as belief … are biological-function categories—very broad and general ones, of course. Compare the categories limb, hormone, circulatory system, eye, visual system, etc. More contentious, the claim includes that such categories or types as belief-that-it-is-raining … are carved out with

[127] The term 'teleofunctionalism' is often used as synonymous with 'teleosemantics', as discussions of teleofunctionalism usually center on the *content* of belief. My discussion centers instead on the teleofunctional individuation of belief (while acknowledging that content plays a role in this individuation). On my use of the term, 'teleofunctionalism about belief' is the view that beliefs are individuated by biological function. This sort of teleofunctionalism is, strictly speaking, independent of teleosemantics—the view that the content of beliefs is specified by biological function—though the two views usually go hand in hand.

reference to biological functions. (1993: 172–173)[128]

In particular, for Millikan, "one of [belief]'s proper functions is to participate in inferences in such a manner as to help produce fulfillment of desires" (71). A believer has the belief-that-it-is-raining insofar as she has a cognitive state that was selected for the dual purposes of affirming the proposition that it is raining and enabling the believer to make useful inferences therefrom.

Paramechanical teleofunctionalism avoids the problems with multiple realizability and intersubjective indeterminacy faced by paramechanical identity theorists, computational functionalists, and pure functionalists. Delia, Roger, and the barista all share the same teleofunctional cognitive state of belief that the mug contains hot coffee insofar as their respective computational (or dispositional) states all serve the same proper purpose. (Each of the café denizens' computational states also serves other purposes, teleofunctionally grounding other beliefs, but their respective cognitive states

[128] Millikan adds a parenthetical: "(though, in the case of beliefs, not directly according to function)" (1993: 173). This is because "being a little more precise, it is the belief-forming mechanisms that produce the adaptations, the adjustments, of the organism to the environment, the beliefs. Beliefs themselves are functionally classified, are "individuated," not directly by function but according to the special conditions corresponding to them that must be met in the world if it is to be possible for them to contribute to proper functioning of the larger system in a historically normal way" (189). So, for Millikan, belief-forming mechanisms and belief-consuming mechanisms are selected for teleofunctions, whereas beliefs are individuated on the basis of their ability—read: systemic capacity function (Cummins 1975)–to aid in the proper functioning of these cognitive mechanisms. This detail of Millikan's teleofunctionalism does not bear on my analysis in the main text, as the signature Rylean move will entail rejecting the notions of attitude-of-belief-forming-mechanisms and attitude-of-belief-consuming-mechanisms alongside the conflation of attitudes and cognitive states of belief. Thus, and for ease of exposition, I will write as if teleofunctionalists assign teleofunctions directly to beliefs.

of belief that the mug contains hot coffee are all individuated by virtue of this shared purpose, abstracting away from implementational differences.) And paramechanical teleofunctionalists can take Ella to have a single, intersubjectively determinate attitude of belief by appealing neither to Ella's current cognitive architecture nor to intersubjective agreement, but to the evolutionary history of the Ella's cognitive state.[129] Jones and Brown both know how Ella actually functions, but the crucial question for the teleofunctionalist is how she properly *ought* to function. If Jones's amateur evolutionary psychology is on the right track, then Ella's belief is anti-suicidal, and Brown is wrong. Otherwise, Ella's attitude of belief is suicidal, and Jones is wrong.[130] Regardless, there is a single intersubjectively determinate fact of the matter about what Ella believes, as determined by the proper biological (rather than actual computational or pure) function of Ella's cognitive state of belief.[131]

---

[129] Dretske (1986) sometimes argues that teleofunctionalism prescribes the ascription of objectively indeterminate beliefs, but that this indeterminacy is harmless. Elsewhere Dretske (1988) goes in for full determinacy. Regardless, what is crucial is that even if there is objective indeterminacy about the teleofunction of belief, this objective indeterminacy could be intersubjectively agreed upon on teleofunctionalist grounds, such that the *intersubjective* indeterminacy dissociation would not apply. In such a circumstance, the teleofunctionalist would hold cognitive states of belief to be objectively indeterminate, and thus have no trouble identifying them with identically objectively indeterminate attitudes of belief.

[130] Of course, it does not take much reflection to find off-kilter the notion that Ella's belief that she ought to kill herself plays a selected-for proper function. Nevertheless, I am going to set aside the notorious problem of maladaptive beliefs, as well as related problems having to do with sophisticated, apparently selectively inert, beliefs (e.g. about quarks and democracy), without comment. For interesting discussions, see Morgan (1883a) and Peacocke (1992), as well as Millikan's (2000a) response to the problem.

[131] Fodor (1992) has influentially argued that teleofunctionalism faces a different kind of indeterminacy worry, which he terms 'the disjunction problem'. In brief, the disjunction problem is that there is no way of saying, for example, whether the frog's belief is that there is a fly there, or that there is a bug there, or that there is a small black object there, etc. Teleofunctionalists often

I previously emphasized dissociations between attitudes of belief and cognitive states analogous to the dissociations between perceivable colors and SSRs. In order to argue against paramechanical teleofunctionalism, I am going to exploit another dimension of the analogy between beliefs and colors. Just as colors co-evolved with organisms' capacities to see color, attitudes of belief co-evolved with organisms' capacities to attribute beliefs. Insofar as etiological teleofunctionalism is a theory of attitudes of belief as well as cognitive states of belief, teleofunctionalists ought to differentiate between the teleofunctions of the former and the teleofunctions of the latter. And insofar as it assigns distinct teleofunctions to attitudes of belief and cognitive states of belief, teleofunctionalism becomes a version of (rather than paramechanical rival to) Ryleanism about belief.

I am therefore not going to argue against teleofunctionalism, though I have strong reservations about its etiological incarnations.[132] Instead, I am going to argue that

---

respond that Fodor's argument only works if teleofunctions are defined internalistically; for example, Sterelny (1990: 125–127) provides an externalistic response to the disjunction problem on behalf of teleofunctionalism. Neander (2017) addresses the disjunction problem with reference to the actual empirical details of toads' fly-detection capacities.

[132] While I do think that philosophers who remain committed to teleofunctionalism should honor the attitude/cognitive state distinction, I do not actually think philosophers should remain committed to etiological teleofunctionalism at all. The foremost of my reservations concerns etiological teleofunctionalists' embrace of natural teleology. For Millikan, for example, beliefs are individuated in terms of their proper functions (or, to be more precise, the ways they contribute to the proper functions of the cognitive mechanisms that consume beliefs). Proper functions are naturally teleological: they invoke the *natural purposes* of the belief in question. I do not think such proper functions exist; I do not think that the process of natural selection imbues adaptations with natural normative purposes (or any other kind of intrinsic normativity) in anything stronger than a metaphorical sense. I thus disagree with those (like Millikan) who think that Darwin made natural teleology naturalistically respectable. On the contrary, I think Darwin's theory obviated naturalistic theorizing that invokes natural teleology; Darwin showed how

teleofunctionalists should give up on their paramechanism: they should uphold the distinction between attitudes of belief and cognitive states of belief while being teleofunctionalists about both. My argument will hinge on questioning paramechanical assumptions about the functional roles attitudes of belief evolved to play, as well as *whose* purposes attitudes evolved to primarily serve. Attitudes of belief do not primarily serve believers; they did not evolve to be cogs in cognitive systems that properly function to help believers transform inputs into outputs and move around their environments. On the contrary, attitudes of belief evolved in tandem with belief attribution capacities, just as perceivable colors evolved in tandem with visual systems.

---

biological entities and processes could be *apparently* teleological without actually being naturally, normatively purpose-driven. As Paul Sheldon Davies writes, "the postulation of such norms puts contemporary theories of functions [including Millikan's theory of proper functions] in conflict with the theory of evolution by natural selection" (2001: *xiv*). See Davies (2001 and 2009) for the arguments that have convinced me that this is the case.

My worry about natural teleology is not the only reason to reject etiological teleofunctionalism as a way of individuating attitudes or cognitive states of belief. For other compelling arguments against teleofunctionalism about the individuation of belief (from philosophers who are largely sympathetic to teleosemantic theories of basic content), see Sober (1985) and Sterelny (1990: 128-137).

Thus, although I will argue in this section that etiological teleofunctionalism can be a kind of Ryleanism, I do not think it provides the best way to account for attitudes of belief (or cognitive states of belief), including their (derived) purposes. My preferred (still broadly teleofunctional) view individuates both attitudes of belief and cognitive states of belief in terms of their respective current systemic capacity functions (Cummins 1975, 1983), rather than etiologically in terms of the proper functions they were selected to possess. Cognitive states of belief function within cognitive systems which cognitive scientists analyze as fulfilling certain (stipulated, non-naturally-normative) purposes. In the case of attitudes of belief, the system in question is the social environment of the mindreader.

Peter Godfrey-Smith (1996) rightly notes that the term 'function' is often used in two senses in biology: a sense that focuses on the current state of the system under analysis and a sense that invokes selective history. But whereas Godfrey-Smith sees these senses as entirely distinct (he thinks the former are Cummins functions and the latter are Wright functions), I see the former as referring to Cummins functions in general and the latter as referring to those Cummins functions that are invoked when biologists explain how a system evolved.

If attitudes of belief have distinctive proper functions, then they properly function within the social environments of belief attributors, as means by which belief attributors are able to regulate, manipulate, predict, explain, and (ethically, aesthetically, epistemically, and pragmatically) evaluate whole patterns of thoughts, feelings, and behaviors.

4.2. Against conflating attitudes with teleofunctional cognitive states

Humans (and many other animals) evolved not only to interact with inanimate objects in their environments, but also to interact with other minded organisms. In Tad Zawidzki's (2013: 233) words, "the most important features of most primate ecologies are social" in a sophisticated sense: they involve primates' recognition of others as end-driven agents. Human social cognition comprises an especially sophisticated set of capacities that were made possible because humans evolved in especially rich social environments.[133] Zawidzki calls this set of capacities the "human sociocognitive syndrome," and takes it to include pervasive cooperation, language use, sophisticated mindshaping ability, and sophisticated mindreading ability.[134]

---

[133] Other proponents of the importance of social environments in human evolution abound. Sterelny writes that "human cognitive competence is a collective achievement and a collective legacy; at any one moment of time, we depend on each other, and over time, we stand on the shoulders not of a few giants but of myriads of ordinary agents who have made and passed on intact the informational resources on which human lives depend" (2012: xii). Those who have been most influential on my thinking (in addition to Sterelny and Zawidzki) include Tomasello (1999), Hrdy (2009), Apperly (2011), Andrews (2012), Chase (2013), and Heyes (2018).

[134] Most theorists countenance these four capacities as (at least nigh) uniquely human. There is, however, no consensus about which of these capacities emerged first in the history of human

Zawidzki calls mindreading "sophisticated" when it involves the explicit attribution of attitudes, as opposed to merely "*some kind of appreciation* of conspecifics' [attitudes] … a kind of sensitivity to or ability to track at least some propositional attitudes" (13).[135] To be able to mindread, in this sophisticated sense, is to be able to attribute attitudes of belief veridically. Though we do not know the precise evolutionary forces that have led to the belief attribution abilities of modern humans, we have good reason to believe that the capacity arose out of the selective pressures of our historical social environments.[136] When you interact with people a lot, it is immensely useful to be able to figure out what they believe. Knowing that a dominant peer does not believe the food you crave exists can help you acquire much needed extra nutrients (Hare et al. 2000). For mothers, reliably attributing beliefs to other adults is crucial to figuring out

---

evolution. Many philosophers take mindreading to have led to the other abilities, but Zawidzki argues that mindshaping is "our sociocognitive linchpin" (2013: 1). Sterelny (2012) and Heyes (2018) rally against the idea of any single magic bullet.

[135] Zawidzki fails to distinguish between attitudes of belief and cognitive states of belief. This leads him to endorse a Dennettian Ryleanism about the targets of unsophisticated mindreading while retaining certain paramechanistic assumptions the attitudes of belief attributed in sophisticated mindreading (see the discussion of two-systems theorists in Chapter 3). In particular, though he stresses that the point of sophisticated mindreading is not to locate the hidden causes of behavior, Zawidzki nevertheless refrains from questioning the paramechanical assumption that attitudes of belief do happen to be "concrete, unobservable causes of behavior" (2013: 11). Needless to say, my use of Zawidzki's account of the evolution of attitudes of belief should not be taken to commit me to all of the details of Zawidzki's (partly paramechanistic) understanding of the nature of attitudes of belief. Nevertheless, we agree about much, including especially the need to reject the paramechanistic assumption that "our quotidian interpretive practices are in the same business as the sciences of the mind, that is, identifying the unobservable causes of behavior" (237).

[136] Heyes (2018) persuasively argues that mindreading (along with the other elements of the human sociocognitive syndrome) may have been the product of cultural rather than genetic evolution. While this topic is fascinating, it does not make a difference to the question of the teleofunctions attitudes of belief (culturally or genetically) evolved to serve.

who to trust to help rear vulnerable, slow-developing human babies (Hrdy 2009). More generally, being able to figure out what weird things our conspecifics believe is often necessary for us to offer satisfactory explanations of their anomalous behavior (Andrews 2012: 224–230). Similarly, attributing certain beliefs to ourselves—and telling our friends what we believe—helps us portray our own behavior as rational and responsible (Malle et al. 2007). Indeed, Zawidzki argues that "our ancestors first started attributing full-blown propositional attitudes … to rehabilitate status in the wake of apparently counternormative behavior, especially apparent reneging on explicit commitments" (224). All in all, the ability to attribute beliefs was plainly adaptive in our ancestral social environments.

Being mindread*able* was adaptive too. If you think, feel, act, and react in patterns that belief attributors can latch onto, then you are more likely to be trusted with precious food, babies, and promises. Developing understandable (and nonthreatening) patterns of dispositions is crucial to being admitted into human societies in which your conspecifics have your back. Crucially, though, these attitudes of belief help believers out only insofar as they render them more intelligible in the eyes of belief attributors (including the believer herself).

These mutually reinforcing adaptive benefits—of the capacity to attribute beliefs and the propensity to live in patterns understandable as beliefs—led to an evolutionary ratchet effect.[137] 'Mindshaping' is a term coined by Matteo Mameli (2001) and developed

[137] This now-mainstream theory of the evolution of social cognition is starkly opposed to the

by Zawidzki.[138] Sophisticated mindshaping comprises a set of practices including imitation, pedagogy, conformity to norms, and narrative self-constitution. What these practices have in common is that they all function to make patterns of behavior more homogenous across a population. In other words, mindshaping delimits the patterns of dispositions to act, react, think, and feel that normal members of a community are likely to have. On Zawidzki's account, sophisticated mindshaping made the evolution of sophisticated mindreading possible. In order to veridically attribute beliefs to other members of their social environments, humans had to first use mindshaping techniques to construct those environments such that beliefs would manifest themselves as understandable—even if sometimes aberrant—patterns of behaviors, thoughts, and feelings. The capacity to attribute beliefs in turn helped people regulate, manipulate, predict, explain, and evaluate the particular thoughts, feelings, and behaviors of their peers on the basis of these patterns, which allowed for ever more sophisticated imitation, pedagogy, conformity, and narrative self-constitution practices. Through this more sophisticated mindshaping, humans developed more refined, more normalized, and thus more predictable, attitudes of belief.[139] Attitudes of belief and mindreading abilities therefore (both biologically and culturally) co-evolved.

Machiavellian Intelligence Hypothesis popularized by Humphrey (1976, 1978). Hrdy (2009), Andrews (2012), Sterelny (2012), and Zawidzki (2013) all discuss compelling evidence against the Machiavellian Intelligence Hypothesis.

[138] Sterelny (2012) terms the same ability "social niche construction."

[139] It might be thought that mindshaping would obviate sophisticated belief attribution, because it would allow for fecund behavior-reading, as suggested by Povinelli & Vonk (2003). I respond to this objection in Chapter 8, which tackles the animal mindreading debate.

Zawidzki's theory of the evolution of social cognition suggests that attitudes of belief were naturally selected in virtue of the functional roles they play in mindreaders' social environments. It was adaptive for our ancestors to be attributed beliefs (and judged favorably on account of those beliefs), just as it was adaptive to attribute beliefs veridically. In Millikan's terms, attitudes of belief have the proper function of enabling belief attribution, which in turn has the proper functions of helping belief attributors predict, explain, regulate, and evaluate the thoughts, feelings, actions, and reactions of believers. To believe that it is raining is to have a pattern of dispositions demarcated by the teleofunction of enabling sophisticated mindreaders to attribute the belief that it is raining to you.

Of course, the natural selection of stable attitudes of belief—qua patterns of dispositions to act, react, think, and feel—inevitably involved the alteration of the cognitive states of believers. Dispositions to act, react, think, and feel are, after all, ontologically dependent on the cognitive systems that enable organisms to act, react, think, and feel. Nevertheless, it is not cognitive states of believers that were selected for as mindshaping practices made it more and more adaptive to behave, think, and feel similarly to everybody else. Successful mindreaders pick up on attitudes of belief qua patterns that fit their models of belief, not qua cognitive states underlying those patterns. It is thus *mindreadable* (rather than purely functional) patterns of dispositions that fulfill the teleofunction of enabling mindreading.

Consider the analogy with the evolution of color perception. There is a lot of

good (though inconclusive) evidence for the hypothesis that trichromacy in primates co-evolved with the colors of the fruits that those primates eat (Regan et al. 2001). It was adaptive for trichromatic primates to more easily locate nutritious fruits, and it was adaptive for colorful-fruit-bearing plants to disperse their seeds via primate digestive systems. Now, this co-evolution inevitably involved the alteration of the intrinsic physical properties and SSRs of the fruits in question. But it is not those physical properties or SSRs that were selected for. The properties of the fruits that were selected for were their perceivable colors, because perceivable colors are the properties that serve the teleofunction of enabling color perceivers to locate fruits (and thereby disperse their seeds).

The moral of this evolutionary fable is that, because colors exist for the purpose of being seen, they cannot be reduced to *invisible* physical and chemical properties of colored objects. Indeed, it is not any non-relational property of fruits, but the way fruits *look* red *to primates* that conferred a selective advantage. Perceivable colors are for color perceivers—they exist in order to be seen by creatures like us—and it is only through being for perceivers that they help plants disperse their seeds.

Analogously, on an account that blends Ryleanism and teleofunctionalism about belief, attitudes of belief are patterns of dispositions to act, react, think, and feel that were selected for the teleofunction of making people recognizable as believers to belief attributors. There are no perceivable colors without (at least historical) color perceivers, and there are no attitudes of belief without (at least historical) belief attributors.

Attitudes of belief are for belief attributors—they exist in order to be grasped by creatures like us—and it is only through being for attributors that they help believers navigate their social environments.

None of this is to deny that there might be cognitive states of belief with the teleofunctions assigned to them by philosophers like Millikan. But it is plausible that if such cognitive states of belief exist, then they acquired their etiological teleofunctions long before belief attribution capacities and attitudes of belief co-evolved. Humans (or their ancestors) had need of cognitive states serving representational and inferential purposes to help them interact with their environments long before they developed sophisticated mindshaping techniques. The two varieties of belief have distinct proper functions, and thus deserve distinct teleofunctional analyses. Whereas cognitive states of belief fulfill (or fail to fulfill) their teleofunctions within the proprietary cognitive systems of believers, attitudes of belief fulfill (or fail to fulfill) their teleofunctions within the social environments of belief attributors. Committed teleofunctionalists should therefore reject paramechanism in favor of a teleofunctional version of Ryleanism about attitudes of belief (while, if they like, retaining their classic teleofunctionalism about cognitive states of belief).[140]

---

[140] It might be objected that attitudes of belief function to give attributors a grip on the cognitive functioning of believers. This is, of course, an instance of Davidson's dogma, but I did not show Davidson's dogma to be false in Chapter 3; I just showed it to be without evidential support. Moreover, I agree that attitudes of belief *can* function, *in part*, to give attributors a loose, indirect grip on the cognitive functioning of believers. But I reject the implication that attitudes of belief can therefore unproblematically be conflated with cognitive states of belief. After all, perceivable colors can function, in part, to give perceivers a grip on how objects reflect light. (Seeing that my

Now, it may be that cognitive states of belief hew extremely closely to attitudes of belief. Indeed, it might be that the evolution of mindreadable attitudes of belief crucially (though contingently) depended on attitudes of belief coming to weakly supervene on cognitive states of belief, such that, given the contingent course of cognitive evolutionary history, believers with the same cognitive states of belief could not have different attitudes of belief. Dretske (2000) offers an argument for the conclusion that beliefs weakly supervene on brain states via an analogy with how monetary value weakly supervenes on the physical properties of coins and bills. Briefly: the social and historical etiology of money guarantees the latter supervenience relation—"thanks to the government's efforts, every piece of paper that has a particular set of intrinsic properties is a genuine $20 bill" (2000: 269). Dretske argues that the evolutionary and developmental etiologies of beliefs analogously guarantee the contingent but complete supervenience of the mind on the brain.

One could run a very similar argument for the weak supervenience of attitudes of belief on cognitive states of belief.[141] Especially if the cognitive states of belief in question were pitched at the personal level (a la Pettit and Jackson 1990) and non-

---

mug is blue gives me a loose, indirect grip on the fact that it reflects more light at the short-wavelength end of the visual spectrum.) Nevertheless, perceivable colors *cannot* be unproblematically conflated with SSRs.

[141] Dretske makes a distinction between behavior and bodily movement, and argues that beliefs cause behavior whereas the brain states they weakly supervene on cause bodily movement. Depending on how it goes, an argument for the weak supervenience of attitudes of belief on cognitive states of belief might entail that whereas cognitive states of belief cause behavior, attitudes of belief render that behavior suitable for folk explanation, prediction, regulation, and evaluation.

etiologically teleofunctionally individuated (a la Sterelny 1990a), this argument might be persuasive. The potential strength of such an argument is what prevents me from arguing that attitudes of belief—or at least the patterns of dispositions they comprise—are necessarily irreducible to cognitive states of belief. Instead, I have argued that attitudes ought not be conflated with cognitive states, given the current level of mystery shrouding human cognitive architecture and the even murkier question of how attributor-relative aspects of mind are coordinated with intrinsic aspects of mind.

Whether or not attitudes of belief weakly supervene on (teleofunctionally individuated) cognitive states of belief is an open empirical question. At the other extreme, eliminativists have long stressed that it is an open empirical question whether cognitive states of belief exist at all.[142] If it turns out that eliminativists are correct that there is nothing to be found in human cognitive systems resembling (much less being supervened upon by) attitudes of belief, then we should be eliminativists about cognitive states of belief. Nevertheless, the attitudes of belief that play functional roles in the social environments of belief attributors will remain ontologically unscathed. If, on the other hand, it turns out that attitudes of belief do weakly supervene on cognitive states of belief, then functionalists will be vindicated in their realism about cognitive states of belief. But they still will not be vindicated in their paramechanism. Weak supervenience does not amount to type-identity (Haugeland 1982). According to

[142] As should go without saying, there are also many empirical possibilities along the continuum between the elimination of cognitive states of belief and the finding that attitudes of belief weakly supervene on cognitive states of belief.

functionalists, the type-identity of any given (attitude or cognitive state of) belief is determined by its function, and attitudes of belief play different functions (in different systems) than cognitive states of belief. Primitive creatures living in a world without belief attributors could have evolved cognitive states of belief, but it would have been impossible for them to evolve attitudes of belief.

Indeed, assigning distinct teleofunctions to attitudes and cognitive states likely restores the double dissociation between the two varieties of belief. Cognitive states of belief that have evolved to serve representational and inferential teleofunctions likely multiply realize attitudes of belief that serve teleofunctions of rendering people recognizable as believers. Moreover, believers with intersubjectively determinate cognitive states of belief can have intersubjectively indeterminate attitudes of belief, since they can have distinct attitudes with teleofunctions serving the distinct purposes of different belief attributors.

Readers who, like me, are uneasy with just so evolutionary explanations of psychological traits need not be alarmed by the adaptationist bent of the preceding discussion. We can see who color and belief are for without reference to evolutionary history. Right now, somewhere in the Blue Ridge Mountains, a berry is being eaten by a bear—leading to the dispersal of its seeds—because of its vibrant color. The functional role of the color is to signal EAT ME! to the bear.[143] Analogously, when we attribute the

[143] Since bears are dichromats, (unlike in the trichromat primate case) it is unclear whether their perceptual systems evolved to discriminate fruits from leaves. Even so, colors plainly do serve bears as relational properties that help them see berries (and other objects they care about.)

belief that the mug contains hot coffee to Delia, Roger, and the barista, we are not speculating about the trio's respective cognitive architectures. We are making sense of, and coming to terms with, their complex patterns of thoughts, feelings, actions, and reactions. The functional role of these patterns is to signal I TAKE THE MUG TO CONTAIN HOT COFFEE! to themselves and other belief attributors.[144] Likewise, the question of what Ella believes matters to Brown and Jones: they are trying to figure out how to intervene, on the basis of whether or not she believes she should kill herself. They are not speculating about what productively causes her to act—by stipulation, they agree about that. Instead, they are arguing about how best to characterize her in terms of belief, so that they can lend a helping hand more effectively.[145]

Teleofunctionalists—etiological or otherwise—ought not conflate attitudes of belief with cognitive states of belief; the former emerge within belief attribution practices, whereas the latter emerge within (etiological or systemic) analyses of cognitive systems. By refusing to conflate attitudes with cognitive states, philosophers can provide a rigorous account of the former that is not vulnerable to revision at the hands of ephemeral trends in theories of cognitive architecture. Mind-brain identity theorists, computational functionalists, commonsense functionalists, teleofunctionalists, and eliminativists about cognitive states of belief can all agree that attitudes of belief are patterns of living that belief attributors identify with taking the world to be some way.

---

[144] See Chapter 6 for more on the social roles played by styles of belief.
[145] See Chapter 7 for more on the social roles played by intersubjectively indeterminate attitudes of belief.

5. Conclusion to Part I

In this chapter and the last, I have offered a preliminary defense of Ryleanism about belief. This defense focused on debunking the paramechanical view that attitudes of belief can be unproblematically conflated with cognitive states of belief. Just as perceivable colors require a distinct philosophical analysis from the light reflectance properties of colored objects, attitudes of belief require a distinct philosophical analysis from cognitive states of belief. The beliefs that people attribute to each other in ordinary life ought to be treated as separate objects of inquiry from the beliefs that cognitive scientists posit as (physical, computational, purely functional, or teleofunctional) cogs in cognitive machines.

5.1. The argument against conflating attitudes with cognitive states

My overarching argument against conflating attitudes of belief with cognitive states of belief proceeds as follows.

P1. In providing an account of a phenomenon, we ought not conflate it with another phenomenon if we have good reason to doubt that the phenomena are type-identical and can provide a satisfactory account of the former phenomenon without conflating it with the latter phenomenon.

P2. We have good reason to doubt that attitudes of belief and cognitive states of belief are type-identical.

P3. We can provide a satisfactory account of attitudes of belief without conflating them with cognitive states of belief.

C. In providing an account of attitudes of belief, we ought not conflate attitudes of belief with cognitive states of belief.

P1 is a general epistemic principle. Since we have good reason to doubt that perceivable colors and SSRs are type-identical and can provide a satisfactory (relationalist) account of the former without conflating it with the latter, we ought to avoid conflating perceivable colors with SSRs. The same reasoning applies to accounts of many psychological phenomena. For example, since we have good reason to doubt that psychogenic seizures and epileptic seizures are type-identical and can provide a satisfactory account of the former without conflating it with the latter, we ought to avoid conflating psychogenic seizures with epileptic seizures.

Chapters 3 and 4 have focused on defending P2. The lack of evidence for the dogma that the folk construe beliefs as inner causes, and the double dissociations between attitudes of belief and brain states, subpersonal computational states, and pure functional states—in addition to the fact that attitudes and cognitive states fulfill distinct teleofunctions—provide ample good reasons to doubt that attitudes of belief and cognitive states of belief are type-identical.

Finally, minimal Ryleanism about belief satisfies P3. The view that beliefs are patterns of dispositions that interpretive schemes designate as constitutive of belief—

unpacked throughout Part I—is a satisfactory account of attitudes of belief that does not conflate them with cognitive states of belief.[146] Indeed, as argued in Chapter 1, Ryleanisms are the only extant realist accounts of attitudes of belief that satisfy P3. Ryleans' exclusive claim to the ability to honor the distinction between attitudes and cognitive states (without denying that the former exist) forms the backbone of my master argument for the truth of minimal Ryleanism about belief.

5.2. The master argument for minimal Ryleanism about belief

The master argument of Part I can be summarized as follows.

MP1. Attitudes of belief are type-identical to whatever sets the veridicality conditions for lay belief attributions.

MP2. Attitudes of belief ought not be conflated with cognitive states of belief.

MP3. If attitudes of belief ought not be conflated with cognitive states of belief, then the veridicality conditions for lay belief attributions are set by the proper objects of lay belief attribution practices.

MP4. If attitudes of belief ought not be conflated with cognitive states of belief, then the proper objects of lay belief attribution practices are patterns of dispositions.

[146] I will continue to defend the viability of Ryleanism as a satisfactory account of attitudes of belief in Part II.

MC. Therefore, attitudes of belief are type-identical to the patterns of dispositions that are the proper objects of lay belief attribution practices.

MP1 states that attitudes of belief are the things (in an ontologically uninformative sense of 'things') that render lay belief attributions accurate or inaccurate. MP1 follows from the uncontroversial definition of 'attitudes of belief' as the beliefs people attribute to one another (and other animals) in everyday life. Even skeptics about attitudes of belief can accept MP1; they simply hold that all lay belief attributions are nonveridical, since people do not really believe anything.

MP2 follows directly from the conclusion of the argument unpacked in the last section, given that I am advancing the master argument in the epistemic context of providing an account of attitudes of belief.

Together, MP3 and MP4 state that MP2 delivers minimal Ryleanism about belief. In particular, MP3 says that the thing that makes lay belief attributions (in)accurate is whatever lay belief attributors properly identify with the beliefs in question, and MP4 says that lay belief attributors properly identify beliefs with patterns of dispositions.[147] Unlike MP1 and MP2, these premises depend on commonsense realism about attitudes of belief: the thesis that Delia, Roger, and the barista must really have *something* in common when they are all aptly said to believe there is hot coffee in the mug. By my

[147] In the jargon established in Chapter 1, P3 asserts that minimal interpretivism follows from P2, and P4 asserts that minimal dispositionalism follows from P2.

lights, commonsense realism is unassailable once we differentiate attitudes from cognitive states, and clarify that realism about attitudes does not entail realism about cognitive states.[148] MP3 and MP4 follow from the combination of commonsense realism and Ryleanism being the only realist game in town that avoids conflating attitudes with cognitive states (as argued in Chapters 1 and 2).[149] There are also positive reasons for thinking the consequents of MP3 and MP4 are true. As argued in this chapter, attitudes of belief play functional roles *for* belief attributors (rather than for believers). And as argued in Chapter 3, Rylean dispositionalism fits the extant empirical evidence about lay belief attribution and explanation snugly, though the jury is still out.

---

[148] If Jackson and Pettit (1990) had embraced the distinction between attitudes and cognitive states, and defended commonsense realism about the former, I would have no quibbles with their approach. Belief attributors obviously successfully latch onto patterns of dispositions—just look at the predictive and explanatory power they unlock! A non-paramechanistic version of Jackson and Pettit's paper would convincingly argue that this predictive and explanatory power suffices to establish realism about attitudes of belief. As Jackson and Pettit argue, Churchland-style paramechanical eliminativism gets its force from the false presupposition that a realism about attitudes of belief must make dubious claims about cognitive architecture.

[149] Here, for example, is the simple only-other-game-in-town hypothetical syllogism that secures the truth of MP3 (that is, the claim that we can derive interpretivism from the nonidentity of attitudes of belief and cognitive states of belief):

IP1. If lay belief attributors are not latching onto cognitive states of belief, then attitudes of belief exist only in relation to lay belief attribution practices.

IP2. If attitudes of belief exist only in relation to lay belief attribution practices, then the veridicality conditions for their attribution are set by the proper objects of lay belief attribution practices.

IPC. So, if lay belief attributors are not latching onto cognitive states of belief, the veridicality conditions for their attribution are set by the proper objects of lay belief attribution practices.

In this respect, my master argument is on similar footing to Davidson's (1963) argument for his dogma, criticized in Chapter 3. Both are arguments from ignorance, and they thus both collapse if an opponent identifies a single viable alternative. The difference is that Ryle provided a viable alternative to Davidson's dogma, whereas nobody has (yet) provided a viable alternative to Ryleanism about attitudes of belief.

Putting MP1–MP4 together, we arrive at a generic sort of Ryleanism about belief: attitudes of belief are constituted by the patterns of dispositions that belief attributors properly identify with taking the world to be some way. However, this master argument delivers only minimal Ryleanism; for one thing, it does not specify which patterns of dispositions attributors properly identify with beliefs. Part II will be dedicated to developing, motivating, and applying the novel, relativistic version of Ryleanism that I find most plausible.

PART II: THE METAPHYSICS OF BELIEF IN PRACTICE

Bellori contemplated everything he saw. Whether it was fish, waterfalls, trees, mountains, birds, insects, or flowers, he saw only the unique. If one reads his notes consecutively, from beginning to end, a feeling is gradually fostered of the infinity of the world. Not "trees" nor even "a tree" but *this* particular tree right here, now, as it is. Not "fish" nor even "a fish" but *this* unique fish right here, now, as it darts suddenly across the sandy bottom through the clear, star-spangled water. Its tail's rapid movement from side to side, the stream of water through the gills, the flat shadow gliding over the bottom beneath it.

– Karl Ove Knausgaard, *A Time for Everything* (2004: 409)

Chapter 5: Beliefs as attributor-relative patterns of living

1. Introduction

Believers emerge in the social environments of people who would categorize them as believers if privy to their latent actions, reactions, thoughts, and feelings. To have an attitude of belief is to live—to be disposed to act, react, think, and feel—in a pattern that an actual belief attributor identifies with taking the world to be some way. By interpreting my niece as believing that her mom loves her unconditionally, for example, I can fit what might be otherwise puzzling individual behaviors—such as freely confessing to her mom that she did something wrong—into a coherent pattern of actions, thoughts, and feelings. Her behavior makes sense given that she is prone to telling her mom about everything important in her life, feeling safe with her mom, thinking that her mom will forgive her, and so on. By attributing a belief to her, I can better understand my niece without making any assumptions about underlying psychological processes or teleofunctional configurations.

I will begin this chapter by noting two wrinkles introduced by my preferred brand of Ryleanism. In the second section, I will further unpack the definition of attitudes of belief with which I opened the chapter. In the third section, I will raise and respond to seven common objections to my definition. This explicatory work will set the stage for Part II: an exploration of how a relativistic Ryleanism about attitudes of belief can illuminate actual practices of belief attribution.

1.1. Two wrinkles in my reformed Ryleanism

The minimal Ryleanism established in Part I is the view that attitudes of belief are type-identical to the patterns of dispositions that are the proper objects of lay belief attribution practices. Minimal Ryleanism does not dictate what counts as a *proper* object of lay belief attribution. Indeed, perhaps the most acrimonious debate between Ryleans concerns the question of how beliefs are fixed in relation to interpretive practices. Nevertheless, Ryleans have traditionally agreed on two tenets beyond minimal Ryleanism. First, Ryleans have agreed that what counts as a proper object of belief attribution does not vary dramatically from believer to believer; if Dave's belief that the PAC-MAN machine is broken consists in a particular pattern of dispositions, then Buster's belief that the PAC-MAN machine is broken will not consist in a dramatically different pattern of dispositions. Second, Ryleans have agreed that what counts as a proper object of belief attribution is determined with reference to a constitutive norm of interpretation, such that what people believe is fixed by either the ideal or normal assessment of their patterns of living.

I reject both of these tenets. Instead of embracing an inflexible dispositionalism and an irreducibly normative interpretivism, my new version of Ryleanism goes beyond minimal Ryleanism by adding the following two elements.

- Beliefs are *stylized* patterns of person-level dispositions.
- Beliefs exist relative to particular models wielded by actual, *individual* attributors.

The thesis that beliefs are *patterns* of dispositions is common among Ryleans, though rarely spelled out in any detail. My reformed Ryleanism adds the wrinkle that the patterns in question often come custom-styled for different believers. Meanwhile, as far as I am aware, I am the only philosopher of mind who commits to the view that beliefs exist only as relativized to *particular* models wielded by *individual* belief attributors. According to my relativistic Ryleanism, the patterns of dispositions that are the proper objects of belief attribution thus vary both from believer to believer and from belief attributor to belief attributor. These idiosyncratic elements of my view set it apart from other Ryleanisms, and provide firepower for the arguments against paramechanical pure functionalism and paramechanical teleofunctionalism pursued in Chapter 4. Other Ryleans would do well to incorporate them, both for their intrinsic merits—which will be extolled in Chapters 6–8—and in order to cleanly dissociate attitudes from cognitive states without invoking the constitutive normativity of the former in a manner that makes many naturalist paramechanists dismissively roll their eyes.

2. What beliefs are (and who they are for)

To believe is to live—to be disposed to act, react, think, and feel—in a pattern that an actual belief attributor identifies with taking the world to be some way.

That is my definition of attitudes of belief. In this section, I will take it apart one piece at a time; I will focus exclusively on explaining the view, rather than defending it. My defense comprises the master argument of Part I (presented at the end of Chapter 4),

the arguments for the two idiosyncratic elements of my Ryleanism in Chapters 6 and 7, and the replies to objections in the third section of this chapter.

2.1. To believe is to live

To believe is to live in the same sense that to have a free throw routine is to play basketball.

First, believers are living creatures, in the same sense that free throw shooters are basketball players.[150] This claim is intended as a generic generalization (Leslie & Lerner 2016); there might be counterexamples and there are certainly borderline cases. Is the 80-year-old podiatrist who set the world record for consecutive free throws a basketball player? Hard to say. It may be similarly hard to say whether certain artificial intelligences are believers, despite not being living creatures. Though I would not be surprised if it turned out to be universally true that believers are living creatures, there is some wiggle-room.

Second, beliefs are characteristics of whole living believers. Elena Delle Donne—who boasts the highest free throw percentage in professional basketball history—has a simple, utterly consistent free throw routine.[151] She dribbles three times, brings her

---

[150] Humans are the paradigm believers, but many other animals are believers too. In Chapter 8, I will defend both the attribution of beliefs to nonhuman animals and the parochial insistence that believing is paradigmatically human.

[151] Demonstrating that consistency, Donne has made 720/765 free throw attempts in the regular season and 82/86 free throw attempts in the playoffs over the course of her WNBA career. That is a conversion rate of 94.24% of her total attempts: almost four full percentage points higher than Steve Nash's NBA-record 90.43%.

elbow up to form a ninety-degree angle, and then bends her knees and shoots in one fluid motion. This free throw routine is a characteristic of Donne as a whole basketball player, as opposed to a characteristic of her elbow or wrist or muscle memory or subpersonal psychological mechanisms of concentration. Now, Donne's elbow, wrist, and cognitive abilities are far from irrelevant to her free throw shooting habits. Sports scientists will want to understand how they contribute. Nevertheless, the free throw routine itself belongs to the whole basketball player. Similarly, my niece believes that her mom loves her unconditionally. This attitude of belief is not a characteristic of my niece's conscious experience, language of thought, prefrontal cortex, or endocrine system. These things are far from irrelevant to her believing. Cognitive scientists will want to understand how they contribute. Nevertheless, the belief itself belongs to the whole living person.

Third, and most centrally, believing is a matter of being positioned to carry out some of the activities (and passivities) of living, just as having a free throw routine is a matter of being positioned to carry out some of the activities of playing basketball. Donne has a rigorous free throw routine even if she is currently asleep: through years of practice, she is positioned to dribble three times before shooting the ball. Analogously, believing is a matter of being positioned to carry out those activities of living associated with taking the belief in question to be true. In virtue of being so positioned, people have beliefs, even while sleeping, whether or not they ever actually carry out the activities associated with believing.

Putting these three aspects together, believing is a variety of living in the Aristotelian sense: it is the whole living creature's first actuality of taking the world to be some way (*De Anima* II:1). To say that somebody believes is to say more than that they have the potential to form a belief that the world is some way (Audi 1994). However, it is not to say that they are occurrently thinking that (or acting as if) the world is some way, though they may well be. Instead, it is to say that they take the world to be some way, whether or not that taking is (ever) made manifest in the here and now.

2.2. To believe is to be disposed to act, react, think, and feel

Translated into the current lingo, this Aristotelian understanding of believing as 'positioning' or a 'first actuality' means that believing consists in having the appropriate dispositions. The appropriate dispositions are tendencies to act, react, think, and feel as if one's belief is true. 'Acting', 'reacting', 'thinking', and 'feeling' all refer to physical and mental occurrences but are otherwise to be construed broadly. Thus, actions include rote habits as well as thoughtful, practically reasoned efforts. Reactions include kneejerks (such as the implicit misogynist's differential bodily postures when listening to men and women speak) as well as carefully honed responses (such as the fencer's parry). Thoughts include occurrent tokens of mental speech as well as unconscious associations. (Thinking does not, however, include believing, insofar as the verb 'believes' picks out the dispositional attitude of belief.) Feelings include everything from tinglings to agonies to tip-of-the-tongue states to enuuis.

The dispositions that compose beliefs are by-and-large thickly described (Ryle 1979; Geertz 1973). My niece's disposition to confide in her mother is not merely a tendency for her muscles to move in various ways. It is a tendency to act, to mean things by her words, to trust her mother, and to allow herself to be emotionally vulnerable. Even the relevant dispositional passivities are aptly thickly described. My niece's disposition to feel safe around her mom cannot be adequately described in merely phenomenological terms, much less merely electro-chemical terms.

Three factors might be built into an adequate thick description of a disposition associated with believing.

First, my niece has a rich mental life, including mental representations and phenomenology, independent of attributors. Despite the hopes of Ryleans like Davidson, Dennett, Baker, and Mölder, I am skeptical that a satisfactory account of original intentionality is going to fall out of a Rylean metaphysics of attitudes of belief. Indeed, I am committed to the view that many mental phenomena exist—and that some mental phenomena are intrinsically representational—independent of practices of interpretation. For example, I am ontologically committed to the representational content of episodes of visual perception; my current visual phenomenology represents the steaming mug on my desk as containing hot coffee. According to my limited, attitude-specific Ryleanism, my niece's belief might involve the disposition to feel warm and comfortable with her mother—and this feeling might constitutively involve a subpersonal mental representation of her mother's love.

Second, the activation conditions of dispositions partly constitutive of attitudes of belief often include other attributor-relative attitudes. For example, my niece's disposition to assert that her mom loves her unconditionally (which partly makes up her belief in the same proposition) is contingent on, inter alia, her attitude of belief that this assertion is socially appropriate in her present context.

Third, many of the manifestations of dispositions might themselves be attributor-relative, insofar as actions, reactions, thoughts, and feelings are almost always presented 'under a description' (Anscombe 1957). Just as Ryleans understand beliefs to exist only as construed by belief attributors, we might understand many individual dispositions—especially dispositions towards *intentional* actions, thoughts, and feelings—to exist only as construed by disposition attributors.[152] Arguments analogous to those provided in Chapter 4 might demonstrate double dissociations between the actions, thoughts, and feelings that concern belief attributors on the one hand, and the muscle movements, mental images, and raw feels undergirding those actions, thoughts, and feelings on the other hand.

---

[152] Among influential Ryleans, Schwitzgebel (2002) and Dennett (1992) deny this third factor; Baker (1995), Davidson (1971), and Ryle (1949, 1979) embrace it. For simplicity's sake, I usually present my view as if this third factor does not obtain: as if the dispositions that constitute relational beliefs are themselves intrinsic properties of believers. (For example, in Chapter 4, I wrote of Brown and Jones agreeing about Ella's dispositions, as if those dispositions were intersubjectively stable intrinsic properties of Ella.) Ultimately, though, my allegiances are with Baker, Davidson, and Ryle. However, my Ryleanism about belief does not depend on this further Ryleanism about actions; I could hold that dispositions are intrinsic properties of believers and still defend my particular brand of Ryleanism about belief.

2.3. To believe is to live in a pattern

Although attitudes of belief are realized by dispositions, they are not identical to mere, unorganized sets of dispositions. Eric Schwitzgebel (2002: 271 fn. 6; 2013: 26) sometimes suggests that belief attributions are shorthand methods for tabulating lists of relevant dispositions. In contrast, I take beliefs to be systematic *patterns* of dispositions. Belief attributors are pattern detectors, and detectable belief-patterns are irreducible to the dispositions they comprise.

The claim that beliefs are real patterns is due to Daniel Dennett (as discussed at some length in Chapter 1). In short, the idea is that the dispositions that make up a belief fit together as more than the sum of their parts: as a (reasonably) unified, coherent way of living as if the world were some way. Often, the pattern itself, as opposed to the pixels that make up the pattern, is the primary object of interest for our various social cognitive practices. When trying to understand my niece, lay belief attributors are rightly disinterested in compiling a complete database of her dispositions. Instead, we are interested in understanding what she believes: in getting a grip on how her patterns of dispositions conspire to make her who she is.

My relativistic Ryleanism emphasizes the stylistic differences between the patterns in which different believers live, even when they believe the very same propositions. Before dwelling on the details of my own account of beliefs qua stylized patterns of living in Chapter 6, I want to stress a commonality between all attitudes of

belief, as agreed upon by all Ryleans. While beliefs are patterns of dispositions, they are not just any old patterns.

2.4. To believe is to live in an identified pattern

Instead, Ryleans hold that beliefs are those real patterns that are picked out as beliefs in lay belief attribution practices.[153]

To constitute a belief, a pattern must be subject to "outer recognition" as a belief, in John Haugeland's sense of the term (1998: 285). In other words, there must be a belief attributor who wields a model of the belief in question, and thus would recognize the pattern of dispositions in question as a belief. For a pattern to be realized as a belief in this manner, it is not necessary that the belief actually be attributed to any particular believer—that is, there need not be (what Haugeland calls) "inner recognition" that somebody's actions, reactions, thoughts, or feelings fit the belief-pattern. It is necessary only that a belief attributor has the outer recognition that if somebody were to live in that pattern, then they would have the belief in question.

Of course, inner recognition and outer recognition frequently go hand-in-hand. As Haugeland notes, the cognitive capacity that enables outer recognition of belief-patterns—the capacity to construct, refine, and store models of belief—evolved partly in

---

[153] This qualification is what distinguishes Ryleanism from pure commonsense functionalism (a la Jackson & Pettit 1990). According to the latter view, beliefs are those real patterns that function to mediate between perceptual inputs and motor outputs, which lay belief attribution practices happen to label 'beliefs'. According to Ryleans, lay belief attribution practices do not merely label; they *transform* patterns into beliefs.

order to facilitate the inner recognition that a believer is behaving in a way that fits the pattern. And the cognitive capacity that enables inner recognition—the capacity to measure evidence about how particular people live against a model, and thereby attribute belief—evolved partly in order to facilitate the outer recognition that belief-patterns exist. Nevertheless, it is important to distinguish the two varieties of pattern recognition, because the distinction reveals how people harbor beliefs even when nobody is occurrently attributing beliefs to them: they live in a manner that a belief attributer outer-recognizes as a belief, even though that belief attributer does not inner-recognize them as a believer. Just as food can be nutritious for an organism whether or not it is actually ingested and metabolized (Hatfield 2003), people can believe for a belief attributor whether or not they are actually inner-recognized as believers.[154] Haugeland's distinction also makes clear how belief attributors can be wrong about what people believe: the attributors might engage in inner *mis*recognition, and thereby misattribute beliefs that they nevertheless aptly outer-recognize to be identical with certain patterns of living. For example, I might (via outer recognition) construe a certain pattern of living as believing that one's mom loves them unconditionally, but (via inner recognition) misattribute belief to my niece because she does not, as a matter of attributor-independent fact, live in the pattern I outer-recognize as believing that her mom loves her unconditionally.

---

[154] Rob Willison (2017: 91) makes a similar point about the outer-recognition of meaning.

How, precisely, to characterize these two cognitive abilities of recognition is underdetermined by the extant empirical evidence. As discussed in Chapter 3, I prefer a model-theoretic account of both outer recognition and inner recognition. The mechanisms in virtue of which models of belief are constructed, refined, stored, and exploited in mindreaders' cognitive systems are the subject of ongoing research and debate in cognitive science.

2.5. To believe is to live in a pattern identified by an actual belief attributor

The most idiosyncratic—and controversial—aspect of my Ryleanism is the view that beliefs are constituted in relation to particular belief attributors' particular models of belief. Strictly speaking, what anybody believes is determined relative to each individual belief attributor's capacity for outer recognition, just as, strictly speaking, the color of any object is determined relative to each individual perceiver's capacity for color experience (Cohen 2010).[155]

This hardline individualistic relativism distinguishes me from Ryleans—like Davidson and Dennett—who take belief to be determined relative to an ideal interpreter or constitutive rational norm of interpretation, as well as Ryleans—like Baker, Mölder, and perhaps Schwitzgebel—who take belief to be determined relative to a normal interpreter or constitutive community norm of interpretation. On my view, we can (and

[155] Recall that the analogy with color is presented for illustrative purposes only; readers need not accept color relationalism in order to find my arguments for relativism about attitudes of belief compelling.

do) generalize from beliefs fixed by individual belief attributors' actual, nonideal outer recognition in order to construct (and discuss, and form policies based on…) notions of an ideal interpretation, or normal interpretation. However, these intersubjective models of belief are abstractions from the metaphysically fundamental beliefs: those that exist relative to the models actually wielded by individual belief attributors. (Of course, at least in the human case, individual belief attributors include the believer herself.)

Arguments for my relativism are forthcoming in Chapter 7, and I reply to the most strident objection—THAT'S RELATIVISM!—in this chapter's last section.

2.6. Beliefs are identified with believers *taking the world to be* some way

Beliefs are not just any old patterns of dispositions that are outer-recognized by actual belief attributors. (Belief attributors outer-recognize all sorts of patterns of dispositions—other attitudes, character traits, and moods, for instance—that are not beliefs.) Instead, beliefs are those patterns that are recognized as *patterns of taking to be* by belief attributors. My niece's belief is identical to the pattern of her dispositions that a belief attributor identifies with her taking her mom to love her unconditionally. Put differently, her belief is her pattern of living *as if* her mom loves her unconditionally. This pattern only includes those tendencies towards actions, thoughts, and feelings that make sense given that the world includes her mom's unconditional love. That is what makes the pattern a belief, rather than a desire, hope, pretense, mood, or personality trait. As discussed in Chapter 3, mindreading researchers have found that the (eventual)

development of the capacity to attribute attitudes of taking the world to be some way (to themselves and other people) is as close to a human universal as psychologists have discovered (Slaughter & Perez-Zapata 2015).

My view is that beliefs are world-directed—"mondial"—and therefore not necessarily propositional (Sommers 2009). My niece's belief is not about a proposition; it is not the mental affirmation that 'that my mom loves me unconditionally' is true (though it may include, as constituent part, the disposition to affirm in mental speech that 'that my mom loves me unconditionally' is true). Instead, my niece's belief is about the world: she takes the world to include her mom's unconditional love for her. I will not defend my mondialism here, however, as nothing else in my view hangs on beliefs being mondial rather than propositional—not even (contra Sommers) the attribution of attitudes of belief to animals (see Chapter 8).

2.7. Beliefs are identified with believers taking the world to be *some way*

I have put it off for four-and-a-half chapters, and I will move past it all too quickly, but here, at last, I am confronted with the problem of intentional content: the problem of what makes a belief that p a belief *that p*. Off the top, I should reiterate that I am a realist about intentionality; many of our mental states—including cognitive states, phenomenal states, and attitudes—are about intentional objects. My niece's belief is about her mother's love, her feeling is about safety, and her confession is about her misbehavior.

I should also reiterate that I am not going to attempt to provide a Rylean account of original intentionality; I doubt that any such account could fall out of an account of attitudes of belief, since attitudes include dispositions to think independently contentful thoughts and utter independently contentful utterances, and exist only in relation to independently contentful models of belief. Dennett, Davidson, Baker, and Mölder are wrong to hold that Ryleanism could account for (or explain away) original intentionality. For my money, original intentionality must be explained either as a byproduct of phenomenal consciousness (Strawson 1994; Horgan & Tienson 2002; Loar 2003; Kriegel 2011), or as arising from the functional configuration of cognitive systems (Fodor 1992; Dretske 1995; Neander 2017). Attributor-relative beliefs thus have, at most, derivative content: derived either from beliefs' constituent dispositions or from belief attributors' models of belief.

The first possibility is that a belief's content is derived from the meaning of propositionally structured manifestations of that belief in outer or inner speech. For example, my niece's belief's content may be derived from the meaning of her dispositional utterance that 'my mom loves me no matter what'. However, not all beliefs have propositionally structured manifestations; consider, for example, the beliefs of animals or small children who lack the capacities for language and propositional thought. This objection does not damn the proposal that the content of beliefs is derived from the content of their constituent dispositions; the content of beliefs may be derived from nonpropositional—perhaps even nonconceptual—content, or nonlinguistic

animals and small children may be second-class believers. Nevertheless, the objection highlights the difficulty of discerning which contents of constituent dispositions are properly responsible for the content of a belief, not to mention explaining the mechanisms by which dispositions imbue beliefs with their contents. For these reasons, I prefer the second possibility.

The second possibility is that a belief's content is derived from belief attributors' ascriptions of content. Belief attributors construe believers as taking the world to be some way or other; perhaps belief attributors thereby imbue patterns of living with intentional content. For example, my niece's belief's content may be derived from my (subpersonal, cognitive model-based) conception of the way in which her belief represents the world (that is, as including her mom's unconditional love).

Ultimately, I am inclined to accept the aspect of the second possibility that dictates that doxastic content is derived from the content ascribed by belief attributors' models. But I am also inclined to reject the dogma that attitudes of beliefs themselves possess content at all. Instead, on my view, content is properly ascribed to whole believers.[156] As Robert Matthews (2007, 2011) has argued—building off work by Dennett, Davidson, and Ruth Barcan Marcus (1990), amongst others—propositional attitude reports measure beliefs by relating believers to content, but this does not mean that beliefs are themselves, strictly speaking, contentful. I am going to be short on this point

[156] Similarly, attitudes of belief are modifications of the character of whole believers, as I will discuss in Chapter 6.

because the (giant) questions of content and the semantics of propositional attitude reports lie beyond the scope of this project. Thankfully, Matthews has already done an excellent job advocating for the measurement-theoretic view.[157] Here is the upshot of Matthews's arguments that I endorse: to report that somebody believes that *p* is not to say that there exists a discrete state of belief that contains a mental representation of *p* obtaining; instead, it is to say that there exists a believer who lives as if *p* obtains. To report that the belief that *p* is true is to say that the world is that way; to report that the belief is false is to say that the world is not that way. My niece lives as if her mom loves her unconditionally; this belief is true because her mom loves her unconditionally. That is all there is to the intentional content that accompanies belief.[158]

3. Seven Common Objections

Putting my definition of the attitude of belief back together again: to believe is to be disposed to act, react, think, and feel in a pattern that an actual belief attributor identifies with taking the world to be some way. My niece lives as if her mother loves her unconditionally, from my (and others') point of view. Per the master argument of

---

[157] As should be obvious from Chapters 3 and 4, while I subscribe to the general thrust of Matthew's measurement-theoretic account of propositional attitude reports, I do not endorse Matthew's (2007) view that beliefs are "aptitudes"—on my view, knowledge may be an aptitude, but belief is merely dispositional—nor the attendant thesis that beliefs causally produce their manifestations.

[158] The claim that content is properly ascribed to whole believers should be distinguished from the minimal Rylean tenet that beliefs are personal (rather than subpersonal) properties of believers. The latter would be true even if beliefs themselves possess content.

Part I and the forthcoming arguments for the two idiosyncratic elements of my Ryleanism, this attributor-relative lifestyle—nothing less and nothing more—is what makes my niece a believer in her mom's love.

In conversation, I have found some responses to my line of thinking to recur time and time again. To conclude this chapter, I will address the seven most common objections.

3.1. Isn't 'taking the world to be some way' just belief? So isn't the definition circular?

Yes, in a reasonable (if not the strictest) sense of 'circular', my definition is circular: it defines beliefs in terms of models of beliefs.[159] But it is not vapidly circular, because the thesis that models of belief metaphysically determine beliefs is a substantive—indeed, hotly contested—claim.

Here is a way of putting my view that makes it sound even more circular, but also shows how it is illuminating: belief is the thing that belief attributors track, when they are tracking how living creatures take the world to be. This arguably circular definition is nonetheless illuminating for several reasons. First, models of belief necessarily construe their targets as takings of the world to be some way, and thus as having properties like content, intentionality, and truth values. This constitutive

[159] Alternately: no, it is not actually circular, because it defines beliefs by way of attributors' models of belief (cf. Searle, 1995: Chapters 2 and 3). This alternative seems like the more apt answer to the objection to me, but interlocutors have insisted that 'beliefs are what models of belief take them to be' is still a kind of circularity, so I give that objection its due in the main text.

construal makes the circle quite wide; belief attributors—insofar as they earn the title—never leave patterns of dispositions entirely unanalyzed. Second, models situate beliefs in relation to other attitudes and characteristics of believers. Belief attributors understand takings the world to be some way to figure in a broader style of living, and attribute them in order to make sense of actions, reactions, thoughts and feelings arising from within a particular life. When belief attributors track how others take the world to be, they thus never track beliefs *alone*; they always track beliefs as components of believers qua whole organisms.

Third, my definition tells researchers that to understand beliefs, we cannot just study believers. We must also study belief attributors, in order to figure out what it is about believers that they track. Fourth, thanks to cognitive scientists' work on theory of mind over the last 40 years, we know a thing or two about belief attribution. We know, for example, that folks across cultures frequently interpret each other as taking the world to be this way or that, and that people attribute these takings in order to rationalize behaviors (among other purposes). Circular or not, my definition shows that by studying the cognitive models that enable belief attribution practices, we shed light on the nature of attitudes of belief themselves.

3.2. Isn't belief attribution itself a kind of belief? So doesn't regress threaten? (Greg's belief is set in relation to Isabel's belief about belief, which is set in relation to

Dave's belief about belief about belief, which is set in relation to Ivy's belief about…)

Yes and no.

Yes, one can (veridically) understand the belief attributor as having a belief about a belief. I believe that my niece believes that her mom loves her unconditionally. This higher-order belief itself exists in relation to another belief attribution, which itself can be felicitously described as a belief. Regress ensues. But this regress is not the least bit theoretically threatening.

The regress is not threatening because it is not the higher-order belief about a belief that outer-recognizes—and thereby metaphysically determines—the pattern of dispositions that constitutes the lower-order belief. Instead, the operative outer recognizer is the cognitive model of belief that is wielded by the belief attributor. This cognitive model is not an attributor-relative attitude. Instead, it is a nonrelational psychological property of the belief attributor, constructed, refined, and applied by her subpersonal social cognitive capacities (as sketched in Chapter 3). So, while we can (and do) talk about belief attribution in terms of a regress of beliefs, beliefs are not metaphysically determined by way of the attitudes of belief that are components of this regress. They are metaphysically determined by belief attributors' cognitive models of the beliefs in question. That is why cognitive scientists who study the subpersonal cognitive mechanisms realizing belief attribution capacities have much to contribute to the project of fleshing out a Rylean metaphysics of attitudes of belief.

3.3. Aren't patterns of dispositions reducible to their categorical bases? So aren't attitudes of belief unproblematically conflated with cognitive states of belief after all?

Individual attitudes of belief are not cleanly reducible to non-attributor-relative properties of believers, because their constituent dispositions feature other attributor-relative attitudes—other beliefs, desires, hopes, worries, and so on—in their activation conditions. Is it possible, then, that the whole holistic web of belief is reducible at once? There are several reasons to suspect not.

A couple of reasons are due to the Rylean U.T. Place. Cognitive states intuitively contribute to causally producing patterns of living, and "as Hume has taught us, [productive] causal relations hold only between distinct existences" (Place 1996: 30). Moreover, people epistemically ascertain cognitive states by taking cognitive systems apart and examining their constituent parts, whereas people ascertain patterns of dispositions by subjecting people to tests. When the activation conditions are fulfilled, does the believer act, react, think, and feel in the manners that they are supposed to be disposed to act, react, think, and feel? If so, then they have the attitude of belief in question. Place reasonably contends that "this epistemological difference is unintelligible on the assumption that both procedures serve to ascertain the existence of the same state" (30). Figuring out what cognitive states somebody has and figuring out what attitudes they have are fundamentally different tasks because cognitive states and attitudes are fundamentally different sorts of state.

I added more reasons to doubt that attitudes reduce to cognitive states in Chapter 4. There are double dissociations between attitudes of belief and any of the cognitive states that are candidates for their categorical bases. Multiple realizability and externalism dissociate attitudes from brain states (4.3.1). Multiple realizability and the fact that beliefs are person-level *patterns* (and we have no reason to assume isomorphic subpersonal patterns) dissociate attitudes from subpersonal computational states (4.3.2). The most promising candidates for categorical bases of attitudes of belief—the candidates that dodge Place's arguments as well as the double dissociations between attitudes and subpersonal states—are functionally individuated nonrelational person-level patterns of dispositions. Nevertheless, multiple realizability and intersubjective indeterminacy dissociate attitudes from person-level functional states individuated by either pure function (4.3.3) or teleofunction (4.4).

Admittedly, it would not be surprising if a mature cognitive architecture revealed that person-level, functionally individuated cognitive states of belief underlie each and every attitude of belief. But this counterfactual lack of surprise does not show that attitudes of belief reduce to cognitive states of belief. It would not have been surprising if it turned out that perceivable colors were type-identical to (much less weakly supervenient on) surface spectral reflectances. Indeed, in my experience, it almost always surprises students that they are not. Attitudes of belief, like colors, arise from the relation between subjects and objects. Maybe they perfectly track cognitive states. But maybe not, since the roles they play in social environments differ from the

roles cognitive states play in cognitive systems. Just as we should not hold the nature and reality of perceivable color hostage to the nature and reality of SSRs, we should not hold the nature and reality of attitudes of belief hostage to the nature and reality of cognitive states of belief.

Whatever one makes of these reasons for resisting reduction, establishing that patterns of dispositions are metaphysically irreducible to their categorical bases is not required in order to defend Ryleanism about attitudes of belief. The debate between Ryleans and paramechanists concerns whether philosophers of belief ought to conflate attitudes of belief with cognitive states of belief. In the context of this debate, folk construals of the objects of lay belief attribution are more pertinent than the metaphysics of dispositions. Even if patterns of dispositions are reducible to their categorical bases—which future philosophers and cognitive scientists may or may not reveal them to be—attitudes of belief play different functional roles in different functional systems than cognitive states of belief. If lay belief attributors attribute beliefs qua patterns of dispositions, without the slightest regard for how believers cognitively function, then philosophers interested in providing an account of attitudes of belief ought not conflate them with cognitive states, the metaphysical issues surrounding dispositions and reduction notwithstanding.

As argued in Chapter 3, we have no good reason to think that lay belief attributors care about cognitive states of belief. If patterns of dispositions outer-recognized by models of belief are the objects of concern in our everyday social practices

of belief attribution, then attitudes of belief are recognizable patterns of dispositions, and ought not be conflated with their categorical bases, whether or not belief-patterns are ultimately reducible to cognitive states. In my view, we *should* be Ryleans about attitudes of belief, *whether or not* we will someday have a compelling story to tell about how Rylean attitudes reduce to teleofunctionally individuated cognitive states, purely functionally individuated personal states, subpersonal computational states, brain states, or all four. And, per the master argument of Part 1, realists about attitudes of belief *must* be Ryleans insofar as there are dissociations between attitudes of belief and all of these candidate cognitive states of belief.

3.4. But aren't belief attributors *really* referring to categorical bases? When I attribute a pattern of dispositions, aren't I always groping for the underlying cause of that pattern, even if that isn't my main pragmatic concern? (Isn't that why Moliere's "dormitive virtue" joke works?)

William Lycan has pioneered an older cousin to this objection. Lycan objects to paramechanical eliminativism about belief on the basis of its implicit descriptivist theory of reference. He argues that if eliminativists were to give up their descriptivism in favor of Lycan's preferred causal-historical theory of reference, then they would also have to give up their eliminativism. On a causal-historical theory of (the variety of) reference going on in belief attribution, Lycan "expect[s] that 'belief' will turn out to refer to some kind of information-bearing inner state of a sentient being, … but the kind of state it

refers to may have only a few of the properties usually attributed to beliefs by common sense" (1988: 32). In other words, Lycan holds that belief attribution involves reference to real subpersonal cognitive states of belief, even if those cognitive states do not much resemble the attitudes of belief captured by lay models.[160]

Objection 3.4—Lycan's objection's little cousin—aims to refute Ryleanism about attitudes of belief, rather than eliminativism about cognitive states of belief, but it works in much the same way. Sure, attitudes of belief might not be type-identical to cognitive states of belief, and belief attributors do not much care about cognitive states of belief. Still, the objection goes, the attribution of 'belief' causally/historically refers to cognitive states of belief, since belief attributors attribute attitudes of belief in order to understand patterns of living, cognitive states of belief causally/historically undergird patterns of living, and, on the causal-historical theory of reference, 'belief' refers to whatever has the right sort of causal/historical link to belief attribution.

This is the most difficult objection to Ryleanism that I have encountered, insofar as I struggle to provide a response that will be compelling to an objector who takes a causal/historical theory of reference as a starting point.[161] It is hard to know what to say in response, beyond "it does not seem to me like I am really referring to the underlying properties of the objects that concern me in the manifest image!" Ultimately, theories of

---

[160] Lycan's argument has been influential: Stich (1997) ultimately gave up his eliminativism in response.

[161] Thanks to Marie Barnett for pushing the objection time and time again, allowing me to try out several varieties of response—none of which Marie has found the least bit compelling.

reference aside, adjudicating this objection comes down to figuring out two things: the purpose of belief attribution, and the purpose of providing a metaphysical account of attitudes of belief.

On Lycan's view, belief attribution is for—serves the final purpose of—getting at the producing causes of behavior (even though, in the end, those causes probably do not much resemble attitudes of belief). On my view, belief attribution is not primarily *for* carrying out flawed attempts at scientifically investigating the producing causes of behavior.[162] Instead, belief attribution is for understanding people and managing social interactions. In understanding people and managing social interactions, attributors are not groping for the producing causes of behavior; they are groping for the recognizable patterns in which people live, whatever their categorical bases. In attributing beliefs, we thus are not referring to the producing causes of behavior, even on a causal-historical theory of reference; instead, we are referring to attitudes of belief themselves. It is only outer-recognized patterns of dispositions, and not their categorical bases, that bear the right sort of causal/historical links to belief attribution.

By way of analogy, consider the folk conception of water: the stuff most humans drink and bathe in daily. Recent work in experimental philosophy shows that lay people identify water with H2O—and not XYZ (Putnam 1975)—in "purely scientific"

---

[162] It can be—and sometimes actually is—*used* for that purpose (Godfrey-Smith 2005). In these cases, it is plausible that belief attribution serves to refer, messily, to whatever the underlying causes of behavior really are. But this contribution to the cognitive scientific enterprise is not among the main roles that belief attribution plays in the social practices that make up ordinary human lives.

theoretical contexts, whereas they identify water with its superficial properties—without regard for its chemical composition—in "legal contexts" with pragmatic upshots (Tobia, Newman & Knobe 2017). These findings indicate that when dealing with the challenges of everyday life, people disregard—and are not really referring to—the deep causal bases of the properties of water that they care about. Instead, lay people refer to the manifest properties of water: its wetness, low viscosity, potability, clean taste, and so on, without any regard for the properties in virtue of which it bears these manifest properties.[163] Likewise, in attributing beliefs to my niece we are not really, even indirectly, groping for her cognitive mechanisms. We are instead referring to how she lives in a recognizable manner, without any regard for the functional and physiological properties in virtue of which she lives in that manner.

These considerations notwithstanding, I begrudgingly admit that it is hard to pull the two referential possibilities apart. As Stich (1997) stresses, any conclusive decision probably depends on the details of one's independently preferred theory of reference as well as various pragmatic considerations. I would add to Stich's list that one's decision about whether belief attributions typically refer to cognitive states of belief crucially depends on one's understanding of what (and who) typical belief

[163] I am not insinuating that water is (or, for that matter, is not) an attributor-relative kind. Instead, I am suggesting that folk conceptions of water in 'legal contexts' and attributions of attitudes of belief are analogous with regard to reference. The point is that, in non-scientific contexts, people rarely refer to the categorical bases of objects; instead, people typically refer to objects' pragmatically relevant surface features—including, especially affordances for action—without (even implicit) regard for substructure.

attribution is for.

I would also add that it depends on what philosophers want out of a metaphysics of attitudes of belief, of the kind developed in these pages. Consider four desiderata for an account of attitudes of belief.

FIGURE 9: FOUR DESIDERATA FOR AN ACCOUNT OF ATTITUDES OF BELIEF

1. An account of attitudes of belief should characterize the actual objects of concern to lay belief attributors in a philosophically satisfying manner.
2. An account of attitudes of belief should accommodate available scientific evidence about attitudes of belief and belief attribution capacities and practices.
3. An account of attitudes of belief should either vindicate realism *or* provide a compelling error theory about the actual objects of concern to lay belief attributors.
4. An account of attitudes of belief should be usefully employable as a theoretical framework for various natural scientific, social scientific, and humanistic investigations about attitudes of belief.

In my estimation, a Ryleanism that distinguishes attitudes from cognitive states—while leaving the door open for either their type-identification or definitive dissociation by future cognitive science—best fulfills these four desiderata.

I fully support speculative theorizing about cognitive states of belief, as well as empirical testing of such theories. Indeed, I think theoretical and empirical work directly addressing the relationships between attitudes, cognitive states, and nervous systems is

a fascinating and promising area of inquiry. Nevertheless, the project of providing a fecund, realist account of attitudes of belief need not be held hostage to future scientific discoveries about cognitive architecture. My relativistic Ryleanism captures the objects of lay belief attribution practices in a manner that is compatible with the full range of cognitive architectures that might realize believers.

The causal-historical theory of reference (as construed by Lycan) implies that a 'surface-level' account that defines attitudes of belief as recognizable patterns of dispositions is explanatorily incomplete. If my four desiderata are on the right track, this implication may be a strike against the causal-historical theory of reference rather than a strike against Ryleanism about attitudes of belief. A paramechanical account according to which attitudes of belief are really mysterious cognitive states which "have only a few of the properties usually attributed to beliefs by common sense" (Lycan 1988: 32) would fail to satisfactorily characterize the objects of concern to lay belief attributors, fail to accommodate the scientific evidence which suggests that belief attributors are unconcerned with inner causes (see Chapter 3), fail to vindicate realism about the objects of concern to lay belief attributors, and fail to be usefully employable as a theoretical framework for the natural, social, and human sciences.

3.5. What about knowledge?

A belief is a pattern of dispositions to act, react, think, and feel. A piece of knowledge is a pattern of capacities: that is, a pattern of dispositions to act, react, think,

and feel *well*. To say somebody believes something is just to say they take the world to be some way, without making any normative judgments about the manner in which they take the world to be that way. Knowledge attributions involve normative judgments. To say somebody knows something is to say that they have the ability to move skillfully through the world in accordance with that thing being the case. In short, Plato got the relationship between belief and knowledge exactly right in the *Meno*. Consult Ryle (1946, 1949, 1979) for more!

3.6. What about desire?

There are two robust folk concepts of 'desire', respectively capturing occurrent urges to have the world be made some way and dispositional wantings that the world be some way. I countenance the former (just as I countenance occurrent tokens of mental speech, which people sometimes refer to as 'beliefs') as non-attributor-relative cognitive states. I countenance the latter as attributor-relative patterns of living that feature occurrent urges as conspicuous manifestations (just like attitudes of belief feature occurrent tokens of mental speech as conspicuous manifestations).

Ryle (1949: Chapter 4) categorized both beliefs and desires (as well as emotions and various other attitudes) as species of "motives": patterns of dispositions towards actions, thoughts, and feelings that serve to explain the purposes of particular manifest actions, thoughts, and feelings. More recently, several philosophers have advocated the view that desire is a subspecies of belief (McNaughton 1988; McDowell 1998; Lauria &

Deonna 2017), and more specifically that a desire is a belief about normative reasons (Gregory 2017). In line with both Ryle's belief-and-desire-as-motives view and this new desire-as-belief view, I am inclined to embrace a general account of attitudes that resists a clean distinction between individual beliefs and desires.

Whereas beliefs are attributor-relative patterns of living as if the world actually *is* some way, desires are attributor-relative patterns of living as if the world *ought* to be some way. Nevertheless, in addition to figuring centrally in the activation conditions of each others' constitutive dispositions, beliefs and desires often share many constitutive dispositions. For example, my niece's tendency to tell her mom secrets partly realizes both her belief that her mom loves her unconditionally and her desire to let the confidential information be known (among many of her other attitudes). I will have more to say about how attitudes like beliefs and desires (as well as habits, emotions, moods, personality traits, and so forth) meld into a comprehensible, holistic conception of the person in Chapter 6.

3.7. That's relativism!

Yes, but…

Yes: beliefs exist only relativized to the models of belief wielded by particular belief attributors. I favor Paul Boghossian's definition of relativism, according to which "the relativist about a given domain, D, purports to have discovered that the truths of D involve an unexpected relation to a parameter," with the rider that "no one of [the

values of this parameter] is more correct for the purposes of determining the facts about [the domain] than any of the others" (2007: 13). My view, as a form of relativism about belief, purports that beliefs exist relative to models of belief; no one of these models determines the facts about belief with more (or less) metaphysical authority than any of the others.[164] In particular, my Rylean relativism about belief is what Boghossian terms a "fictionalist" brand of relativism. "Compare truths about fictional characters:" Boghossian instructs, "there are no truths of the form *Sherlock Holmes lived on Baker Street*, but only ones of the form *According to the stories of Arthur Conan Doyle, Sherlock Holmes lived on Baker Street*" (2007: 24). On my view, there are no truths of the form *Devin's niece believes that her mom loves her unconditionally*, but only truths of the form *According to Devin's cognitive model of the belief that one's mom loves them unconditionally, his niece believes that her mom loves her unconditionally*.

But: fictionalist relativism is ill-named insofar as it does not include the claim that the properties in its domain are fictional. It dictates only that the properties in its domain exist relative to some parameter, just as fictional characters exist relative to

---

[164] Further clarifications: my individual attributor relativism is neither contextualism not assessment-sensitivity relativism. First of all, unlike contextualism, Ryleanism is a view about beliefs themselves, not a view about the content of sentences (or propositions) about beliefs. As Boghossian (2006: 15–16) would point out, belief attributors don't intend their apparently non-relational belief-attribution sentences—e.g., 'Paul believes that the sky is blue'—to be elliptical for relational sentences—e.g., in relation to Devin's model of belief, Paul believes that the sky is blue'. Moreover, my version of Ryleanism holds that what somebody believes depends on belief attributors' models of belief, not necessarily on the context of actual belief attribution nor the context of the assessment of actual belief attribution. It is thus not a version of relativism according to Macfarlane's (2014: 24) definition of relativism as assessment-sensitivity.

works of fiction. Thus, my Rylean relativism denies neither that beliefs are real nor that there are objective—and objectively assessable—facts about what people believe (just as Galilean relativism about motion denies neither that motion is real nor that there are objective—and objectively assessable—facts about how objects move). Instead, I deny only that there is any single, privileged, and absolute fact about what somebody believes (just as Galileo denied that there is any single, privileged, and absolute fact about how objects move). Instead, there are (at least) as many facts about what somebody believes as there are belief attributors (just as there are as many facts about how objects move as there are frames of reference). Each of these facts is perfectly objective: there is a single true account of what my niece believes in relation to my capacity for belief attribution, another single true account of what my niece believes in relation to her own capacity for belief attribution, and a third true account of what she believes in relation to Galileo's capacity for belief attribution.

And: we can and do talk about—and build social practices on the basis of—beliefs in terms of intersubjective models of beliefs, qua abstract generalizations from particular belief attributors' models of belief. For example, when news anchors invoke 'racism', they mean (something like) 'the belief that the members of one race are inferior to the members of another race in virtue of their race'. But they do not mean to invoke that belief just as it exists in relation to the news anchor themselves, much less just in relation to any particular cable TV viewer. They mean to invoke the racist belief as it is generally modeled by the population at large. My relativism about belief does not deny

the viability of this kind of folk discourse about racism, or other beliefs-relativized-to-intersubjective-models. Instead, my view is simply that this intersubjective kind of "belief" is nothing more and nothing less than an abstract generalization—and, appropriately, Galilean idealization (Weisberg 2013)—from racism as it exists relative to the news anchor, each and every viewer, and everybody else in the target population (maybe discounting outliers).

Abstract intersubjective models of belief are fairly reliable, and folk discourse about beliefs functions fairly smoothly, because individual humans' models of belief do not differ all that much from belief attributor to belief attributor. Despite four hundred years of temporal distance, Galileo and I would have no trouble agreeing on what my niece believes (given equal evidence of her behavior, thoughts, and feelings) because, contingently but not at all surprisingly, Galileo and I wield extremely similar models of many beliefs. Several intense causal pressures interact to ensure that folks—especially folks who share a culture—do not differ much in how they conceive of beliefs.

Consider two of the most salient of these pressures: evolution has naturally selected for belief attributors who can understand believers in normal ways (as discussed in Chapter 4.4), and societal models of belief shape individual models of belief. Regarding the latter, consider a child who is learning to become a mature belief attributor at her parents' knees. The adults will "correct" their child when she "misattributes" beliefs—that is, when she applies models of beliefs that do not match her parents'—leading the child to alter her models to better fit her parents'. Through social

processes of mindshaping, such as the child being taught norms of belief attribution by her parents, models of belief become (more or less) normalized across populations. And insofar as models of belief become normalized across populations, intersubjective models of belief (which are constructed by generalizing from individual models of belief in a population) become better equipped to track actual, individual attributor-relative models of belief more and more faithfully.

Still: attitudes of belief are inextricably relativized to the models wielded by particular belief attributors. This aspect of my Ryleanism runs counter to the views of all other Ryleans of whom I am aware (not to mention all the non-Ryleans). As mentioned in Chapter 1, Schwitzgebel (2002: 271, fn. 6.) might be the lone exception, although his flirtations with relativism are relegated to footnotes and immediately downplayed therein. Schwitzgebel's (in draft) recent pragmatist approach to the metaphysics of belief has led him to drift even further from his youthful flirtations with relativism. Unlike Schwitzgebel, I let my relativism flag fly; I think relativism about the attitude of belief is true, and that it is not a truth that Ryleans should be embarrassed to admit.[165] Rylean relativism about belief is, in the main, a narrowly metaphysical stance: it is the view that

---

[165] It is worth mentioning that even if my Rylean relativism about attitudes of belief is on the money, it is possible that attitudes are weakly supervenient on (though not type-identical to) cognitive states, insofar as a mature cognitive science might recognize many viable ways of carving up cognitive systems. Thus, Ella might really have different attitudes of belief in relation to Brown and Jones, even though her attitude-of-belief-for-Brown and her attitude-of-belief-for-Jones are both weakly supervenient on cognitive states of belief. (Depending on how the theory of cognitive architecture I am imagining goes, the two cognitive states in question may themselves only be visible relative to incompatible systemic capacity analyses of Ella's cognitive system.)

attitudes of belief are metaphysically constituted in relation to particular belief attributors' models of those beliefs, but it carries the rider that we can unproblematically carry out most of our practical affairs as if there were absolute, non-attributor-relative facts about what folks believe. Nevertheless, as I will argue in Chapter 7, the real phenomenon of conflicting models of belief can sometimes make a practical difference in everyday social practices involving belief attribution and, as I will argue in Chapters 7.5 and 8, countenancing relativism about belief can lead to methodological improvements in the natural and social sciences investigating attitudes of belief.

4. Conclusion

So much for the blueprint for my view and preliminary replies to objections. Time to get down to the nitty gritty details. Chapter 6 will delve deeper into the stylized belief-patterns that lie at the heart of the dispositionalist aspect of my Ryleanism. Chapter 7 will defend the relativistic interpretivist aspect of my Ryleanism. Chapter 8 will apply Ryleanism about belief to both substantive and methodological questions in animal cognition research.

Chapter 6: Styles of belief

1. Introduction

Human belief attribution practices are diverse, complex, and flexible. In this chapter and the next, I develop my Rylean metaphysics of attitudes of belief by exploring the sometimes stark, sometimes nebulous differences between the beliefs that people veridically ascribe to one another. In Chapter 6, I will show how Ryleans can account for variance in styles of belief between believers. This chapter primarily concerns the dispositionalist element of Ryleanism, and, in particular, the question of how to deal with diversity in the patterns of dispositions that make up believers' shared beliefs. In Chapter 7, I will show how Ryleans, and Ryleans alone, can account for variance in the conditions of veridical belief attribution between belief attributors. Chapter 7 primarily concerns the interpretivist element of Ryleanism, and, in particular, the question of how to deal with diversity in how belief attributors model beliefs. Together, these chapters flesh out the claims about styles of belief and intersubjective indeterminacy that powered the master argument for Ryleanism presented in Chapter 4, and in so doing combine a new dispositionalism with a new interpretivism to produce a new Ryleanism.

1.1. Ryle on Austen

There is an apocryphal story about a journalist asking Gilbert Ryle if he ever

found the spare time to read novels. Ryle reputedly responded: "Oh yes. All six, every year."

Contrary to the journalist's suggestion, Ryle's time spent reading the books of his favorite novelist, Jane Austen, was not time spent away from doing philosophy. Ryle found philosophical insights to permeate Austen's works. In the last chapter of *The Concept of Mind*, Ryle brought these insights up in the midst of a discussion about what he thought the behaviorists got right about human psychology.

> Novelists, dramatists and biographers had always been satisfied to exhibit people's motives, thoughts, perturbations and habits by describing their doings, sayings, and imaginings, their grimaces, gestures and tones of voice. In concentrating on what Jane Austen concentrated on, [behaviorist] psychologists began to find that these were, after all, the stuff and not the mere trappings of their subjects. (Ryle 1949: 328)

To my ear, this is the clearest short statement of what Ryleanism has in common with behaviorism. Like the behaviorists, Ryle resolutely refused to identify the mental—people's motives, thoughts, perturbations, and habits—with anything underlying the thick, personal level which fascinated Jane Austen.

In a later essay, Ryle asserted that Austen wrote "partly from a deep interest in some perfectly general, even theoretical questions about human nature and human

conduct" (1966: 286). "She wrote *Pride and Prejudice*," for example, "from an interest in the quite general question what sorts and degrees of pride do, and what sorts and degrees of pride do not go with right thinking and right acting" (290). Ryle argued that Austen's novels delivered properly Aristotelian answers to such questions. However, Ryle commended Austen for refraining from doing philosophy in the style of an Aristotelian philosopher. Aristotelians like Theophrastus took such questions one at a time, and in the abstract. What is it to believe that a single man in possession of a good fortune must be in want of a wife? Theophrastus would respond by reflecting on that belief alone, abstracted from the particular ways in which different believers hold the belief, so as to provide a candidate general analysis. But this was not Jane Austen's way. "In contrast," Ryle wrote that

> Jane Austen's technique is the method of the vintner. She pin-points the exact quality of character in which she is interested, and the exact degree of that quality, by matching it against the same quality in different degrees, against simulations of that quality, against deficiencies of it and against qualities which, though different, are brothers or cousins of that selected quality … To discriminate the individual taste of any one character is to discriminate by comparison the individual taste of every other character. That is to say, in a given novel Jane Austen's characters are not merely blankly different, as Cheltenham is blankly different from Helvellyn. They are different inside the same genus, as

> Cheltenham is different from Bath or Middlesbrough, or as Helvellyn is different from Skiddaw or Boar's Hill. (1966: 288)

Ryle went on to specify how the characters in *Pride and Prejudice* are different inside the same genus: all proud, but possessed of diverse styles of pride.

> Thus in *Pride and Prejudice* almost every character exhibits too much or too little pride, pride of a bad or silly sort or pride of a good sort, sham pride or genuine pride and so forth. Elizabeth Bennet combines a dangerous cocksureness in her assessments of people with a proper sense of her own worth. Jane is quite uncocksure. She is too diffident. She does not resent being put upon or even realise that she is being put upon. There is no proper pride, and so no fight in her. Their mother is so stupid and vulgar that she has no sense of dignity at all, only silly vanities about her dishes and her daughters' conquests. Mr Bennet has genuine pride. He does despise the despicable. But it is inert, unexecutive pride. He voices his just contempt in witty words, but he does nothing to prevent or repair what he condemns. It is the pride of a mere don, though a good don. Bingley has no special pride, and so, though a nice man, spinelessly lets himself be managed by others where he should not. His sisters are proud in the sense of being vain and snobbish. (289)

Ryle continued with the assertion that

> Darcy is, to start with, haughty and snobbish, a true nephew of Lady Catherine de Burgh. His early love for Elizabeth is vitiated with condescension. He reforms into a man with pride of the right sort. He is proud to be able to help Elizabeth and her socially embarrassing family. He now knows what is due from him as well as what is due to him. Mr Collins is the incarnation of vacuous complacency. He glories in what are mere reflections from the rank of his titled patroness and from his own status as a clergyman. He is a soap-bubble with nothing at all inside him and only bulging refractions from other things on his rotund surface. (289)

Pride and Prejudice imparts no general analysis of either pride or prejudice, but Ryle argued that it does, nevertheless, provide the reader with theoretical insight into the nature of these character traits. In particular, Austen shows us that there are many degrees and varieties of pride, and that these varieties interact in subtle—albeit sometimes extravagantly comedic and dramatic—ways. Merely to attribute pride to someone is to explain very little about them. However, to attribute genuine but inert, unexecutive pride to Mr. Bennett explains much about what sort of man Elizabeth Bennett has for a father.

In *The Concept of Mind*, Ryle put Austen's work to explicit theoretical use. Ryle

characterized mental states like beliefs as "dispositions the exercises of which are indefinitely heterogeneous" (1949: 44): that is, patterns of dispositions. In order to motivate this thesis, Ryle again invoked Austen's treatment of pride in *Pride and Prejudice*.

> When Jane Austen wished to show the specific kind of pride which characterised the heroine of 'Pride and Prejudice', she had to represent her actions, words, thoughts and feelings in a thousand different situations. There is no one standard type of action or reaction such that Jane Austen could say 'My heroine's kind of pride was just the tendency to do this, whenever a situation of that sort arose'. (1949: 44)

In the next paragraph, Ryle argued that this goes for belief as much as pride.

> … Flouting the testimony of experience, [epistemologists] postulate that ... a man who believes that the earth is round must from time to time be going through some unique proceeding of cognising, 'judging', or internally re-asserting, with a feeling of confidence, "The earth is round'. In fact, of course, people do not harp on statements in this way, and even if they did do so and even if we knew that they did, we still should not be satisfied that they believed that the earth was round, unless we also found them inferring, imagining, saying and doing a great

> number of other things as well. If we found them inferring, imagining, saying and doing these other things, we should be satisfied that they believed the earth to be round, even if we had the best reasons for thinking that they never internally harped on the original statement at all. (44)

The main point of this passage is that (inner or outer) assertions that *p* are neither necessary nor sufficient for belief that *p*. But there is also an implicit, secondary point: believing that *p* is *structurally* analogous to being prideful. Ryle meant two things when he said that beliefs, like pride, are "multi-track"—or "determinable"—dispositions, "the exercises of which are indefinitely heterogeneous." First, and most famously, he meant to refute what is now termed "intellectualism" about belief (Marley-Payne 2016; Schwitzgebel in draft). Ryle stressed that a belief does not consist *solely* in the "one-track intellectual process" of being disposed to assent to a proposition. Second, and equally importantly, he suggested that different people can have the same belief *in different ways*. Just as most of the characters in *Pride and Prejudice* are prideful, but prideful in different ways, so too do most believers that *p* have different styles of believing that *p*. A theory of attitudes of belief—especially a Rylean theory of attitudes of belief—must have the tools to account for how believers can be "different inside the same genus": how they can be detected to match the same pattern in different respects.

In this chapter, I am going to provide those tools.

2. Ryleans on variation between believers

Several Rylean philosophers have remarked on the fact that different people can have the same belief in different ways.

2.1. Dennett and Baker

Daniel Dennett describes such a case as follows.

> Jacques shoots his uncle dead in Trafalgar Square and is apprehended on the spot by Sherlock; Tom reads about it in the *Guardian* and Boris learns of it in the *Prada*. Now Jacques, Sherlock, Tom, and Boris have had remarkably different experiences—to say nothing of their earlier biographies and future prospects—but there is one thing they all share: they all believe that a Frenchman has committed murder in Trafalgar Square. They did not all *say* this, not even "to themselves"; *that proposition* did not, we can suppose, "occur to" any of them, and even if it had, it would have had entirely different import for Jacques, Sherlock, Tom and Boris. Yet they all believe that a Frenchman committed murder in Trafalgar Square … Ordinary folk psychologists have no difficulty imputing such useful but elusive commonalities to people. (1987: 54–55)

Jacques, Sherlock, Tom, and Boris diverge sharply in many of the dispositions that make up their respective beliefs that a Frenchman committed murder in Trafalgar Square.

Nevertheless, their divergent ways of believing share a common core of dispositions, and these shared dispositions are enough to qualify them as all having the same belief. Dennett writes that

> … it is their sharing of *this* belief that would explain (or permit us to predict) in some imagined circumstances their all taking the same action when given the same new information. ("And now for one million dollars, Tom [Jacques, Sherlock, Boris], answer our jackpot question correctly: has a French citizen ever committed a major crime in London?") (56–57)

Lynne Rudder Baker takes this Rylean embrace of diversity of dispositions that make up the same belief to its extreme, claiming that "'x believes that p' may be true of S and S' even if the relevant counterfactuals for S and S' have few, *if any*, members in common" (1995: 156–157).

Unfortunately, Ryleans have tended, as a rule, to avoid providing many details about what people that believe that *p* in diverse manners all have in common. Dennett simply notes that the commonalities are elusive, while insisting that they must exist, since folk psychologists make such good use of them. Baker, meanwhile, admits the possibility that two believers that *p* have absolutely nothing in common—not even the disposition to answer '*p*!' for a million dollars—other than the "irreducible fact" that they both believe that *p* (1995: 157).

2.2. Schwitzgebel

The one exception to this rule is Eric Schwitzgebel (2001, 2002, 2010, 2013). Like Ryle, Schwitzgebel explicitly analogizes beliefs to character traits. To be brave is to be disposed to act in ways that people stereotypically associate with bravery. Likewise, Schwitzgebel claims that "to believe that P … is nothing more than to match to an appropriate degree and in appropriate respects the dispositional stereotype for believing that P" (2002: 251).

The key notion here is that of a "dispositional stereotype." Schwitzgebel defines a stereotype, with reference to the literature in psychology, as "a cluster of properties we are apt to associate with a thing, a class of things, or a property" (250). Thus, a *dispositional* stereotype *for believing that p* is a cluster of dispositional properties that an interpreter would be apt to associate with the belief that *p* (251). Schwitzgebel writes that dispositional properties "can be characterized by means of conditional statements of the form: If condition C holds, then object O will (or is likely to) enter (or remain in) state S" (252). The dispositional properties that Schwitzgebel takes normal interpreters to be apt to associate with beliefs include not only dispositions to engage in certain behaviors, but also "*phenomenal* dispositions, dispositions to have certain sorts of conscious experiences," and *cognitive* dispositions, "dispositions to enter mental states that are not wholly characterizable phenomenally" (252). Putting these definitions together, the dispositional stereotype for believing that *p* is the cluster of behavioral, phenomenal, and cognitive dispositions that a normal interpreter would be apt to associate therewith.

Schwitzgebel's 'dispositional stereotypes' are thus cognitive models of belief with built-in dispositionalist construals.[166]

Consider the belief that a Frenchman committed murder in Trafalgar Square. On Schwitzgebel's account, this belief would be defined with respect to a dispositional stereotype that might include the behavioral disposition to assert, when asked if there is any news, that there was a murder in Trafalgar Square, the phenomenal disposition to feel nervous when meeting a Frenchman in Trafalgar Square, the cognitive disposition to infer that a Frenchman committed a major crime in London, and so on.

Now, in order to believe that a Frenchman committed murder in Trafalgar Square, Jacques, Sherlock, Tom, and Boris need not exhibit, or even come close to exhibiting, any particular manifestations of the dispositions that make up the dispositional stereotype for that belief. Some of the activation conditions required for the manifestation of behavioral, phenomenal, and cognitive dispositions are, after all, conditions (in both the logical and psychological senses of the term) *of the believer*. Jacques will probably only ever assert that a Frenchman committed murder in Trafalgar Square if he desires to make that assertion. For obvious reasons, Jacques might never form that desire. Sherlock, for his part, might never feel nervous among Frenchmen in Trafalgar Square, given his even-keeled frame of mind. Finally, if Boris does not know that Trafalgar Square is in London, then he will not make the inference that a major

---

[166] See Chapter 1.2.3 for more on Schwitzgebel's Ryleanism, Chapter 3.3 for more on cognitive models and their construals, and Section 3.4 of this chapter for a way of more explicitly combining Schwitzgebelian Ryleanism with model-theoretic views of folk psychology.

crime occurred in London.

In some of these cases it is hard to say whether (on the one hand) the believer lacks the disposition in question, or whether (on the other hand) the believer has the disposition but the activation conditions never obtain. I will offer a heuristic for dissolving this ambiguity shortly. In the meantime, note that this ambiguity does not necessarily pose a serious problem for Schwitzgebel's account of belief. For in order to believe that a Frenchman committed murder in Trafalgar Square, believers need not possess, much less manifest, *all* of the dispositions in the dispositional stereotype for that belief. As Schwitzgebel notes, believers that *p* only have to match the dispositional stereotype for belief that *p* "to an appropriate degree and in appropriate respects." Schwitzgebel argues that "what respects and degrees of match are to count as 'appropriate' will vary contextually and so be left to the ascriber's judgement" (253). On Schwitzgebel's account, the dispositional stereotype for any given belief does not vary with the context of ascription. Instead, what vary are the respects and degrees of match with the stable dispositional stereotype that legitimize ascription of belief.

2.3. Juliet

To illustrate his theory, Schwitzgebel describes a series of cases of "in-between belief." In the case that has received the most attention in the literature, Schwitzgebel tells the story of Juliet, an implicitly racist academic who "has critically examined the literature on racial differences in intelligence, and … finds the case for racial equality

compelling," yet "is systematically racist in most of her spontaneous reactions, her unguarded behavior, and her judgments about particular cases." Juliet "is prepared to argue coherently, sincerely, and vehemently for equality of intelligence and has argued the point repeatedly in the past." However,

> when she gazes out on class the first day of each term, she can't help but think that some students look brighter than others—and to her, the black students never look bright. When a black student makes an insightful comment or submits an excellent essay, she feels more surprise than she would were a white or Asian student to do so, even though her black students make insightful comments and submit excellent essays at the same rate as do the others. This bias affects her grading and the way she guides class discussion. (2008: 532)

The question is: does Juliet believe black people are intellectually inferior to white people? Paramechanist philosophers of mind have offered (at least) four different answers to this question. First, perhaps Juliet believes black people are intellectually inferior to white people, but has a higher-order judgement that she ought not believe it (Hunter 2011). Second, perhaps she disbelieves it, but has another, different mental state (an alief, for example) that causes her racist behaviors (Gendler 2008). Third, perhaps she has contradictory beliefs, and both believes and disbelieves that black people are intellectually inferior (Sommers 2009; Mandelbaum 2013). Finally, perhaps she shifts

back and forth between states of belief and disbelief (Rowbottom 2016).

Schwitzgebel has argued against all of these ways of understanding how Juliet believes. He points out that belief is "a continuous phenomenon" (2002: 249), not an all-or-nothing affair. It seems wrong to ascribe to Juliet either the same racist belief as a white supremacist or the same anti-racist belief as her fellow activists, whose behavior more consistently matches their avowed beliefs in intellectual equality. Instead, Schwitzgebel diagnoses Juliet as being in a state of "in-between believing". What does Juliet really believe? Schwitzgebel argues that paramechanists would have to decide: at any given moment, she either believes that black people are intellectually inferior to white people or she lacks that belief entirely.[167] But neither of these answers seems satisfactory. So Schwitzgebel points out that on a Rylean theory of belief, we do not have to choose between them.

Here is the messy fact of the matter: Juliet matches the dispositional stereotype for believing that black people are intellectually inferior to white people to some degree and in some respects, and she matches the dispositional stereotype for believing that all races are intellectually equal to some degree and in other respects. Juliet lives in-between racism and anti-racism. In some contexts of ascription, it is appropriate to ascribe to Juliet the belief that black people are intellectually inferior. For example, it

[167] Paramechanists have rightly pushed back against this argument of Schwitzgebel's. For example, Quilty-Dunn & Mandelbaum (2017) argue that cognitive states of belief might be stored in fragments, such that Juliet's in-between belief is explainable in terms of her having both beliefs, store separately and activated by distinct cognitive processes.

might be appropriate to ascribe this belief to her when conducting a review of her teaching and grading practices. In other contexts—while she is pontificating about the literature on racial differences and intelligence, for instance—it might not be appropriate to ascribe that belief to her. Finally, considered in general, abstracted from any particular context of ascription, the only determinate answer to the question 'does Juliet believe that black people are intellectually inferior to white people?' is 'sort of'.

3. Improving on Schwitzgebel's conceptual machinery

Schwitzgebel's discussion of in-between belief supplies a solid, though by no means decisive, argument for Ryleanism. However, Schwitzgebel's major contribution is the conceptual machinery he introduces. The notion of dispositional stereotypes for belief puts flesh on the bones of Ryleanism, even though, as Schwitzgebel admits, "it does not give a simple, straightforward criterion for assessing the presence or absence of belief" (269). As such, I am going to build my relativistic version of Ryleanism from Schwitzgebelian building blocks.

One way to improve upon Schwitzgebel's brand of Ryleanism might be to go into great detail about the structure of the dispositional stereotypes we wield for certain beliefs. For example, Schwitzgebel insinuates that dispositional stereotypes for belief are not best conceived of as mere lists of dispositions, but as more finely structured clusters of dispositions (2013: 9). Some dispositions are more central to the stereotype for any given belief than others. The disposition to sincerely assert *p* (when asked whether *p* by

a confidant) is probably central to the dispositional stereotypes of most beliefs that *p*. Thus, we might ask which (kinds of) dispositions tend to be central to the dispositional stereotypes of certain classes of beliefs, and which (kinds of) dispositions tend to be more peripheral. Or we might ask how we can articulate the threshold at which it makes sense to say (in a certain context) that Juliet *does* believe that all races are intellectually equal, though she might not believe it as thoroughly as her colleagues.

Answers to these questions would be invaluable. To be frank, however, I am not optimistic about the success of endeavors to dramatically clean up Schwitzgebel's conceptual machinery. For one thing, any such endeavor would have to rely on a huge amount of experimental and ethnographic data about how diverse people attribute particular beliefs in diverse contexts.[168] Any conceptual machinery that could do a reasonably good job at articulating the metaphysics of attitudes of belief would have to be messy and complex, because, with Schwitzgebel, I think the practices of belief ascription that bring attitudes of belief into being are themselves messy and complex.

In what follows, I am going to offer a friendly amendment to Schwitzgebel's account that makes his machinery more, rather than less messy and complex. Or (if you are feeling charitable) I am going to improve on Schwitzgebel's machinery by making it richer and more flexible, so as to better account for the richness and flexibility of our folk practices of belief ascription. In particular, I am going to suggest not only that believers

[168] I discuss the fledgling empirical literature that exists on this topic in Chapter 7.

differ in degree and variation of fit to general-purpose dispositional stereotypes, but also that belief attributors often wield strikingly different believer-specific models[169] when assessing different believers.[170]

3.1. Cases of variance between believers

In order to deal with cases of dissonance between professed beliefs and behavior, Schwitzgebel treats beliefs like character traits. Just as the atheist in the foxhole can be sort of brave—that is, can match a model for bravery to certain incomplete degrees and in certain incomplete respects—she can sort of believe in God. But as we saw in my opening bit about Ryle on Austen, character traits have another, related feature. There are degrees of pride. But there are also different *styles* of pride. In order to complete the analogy with character traits, I am going to have to account not only for in-between beliefs, but also for the wide array of styles in which different people—who all fully believe that *p*—think, feel, and behave as if *p*. Consider a couple cases.

Case #1: Atticus impeached

Atticus Finch, American man of integrity, has been impeached. He is a racist.[171]

---

[169] From here on out, I use the term 'model of belief' (Maibom 2003; Godfrey-Smith 2005) interchangeably with Schwitzgebel's 'dispositional stereotypes'. I tend to use 'stereotype' instead of 'model' only to refer to general-purpose models—which rely on gross generalizations more than believer-specific models (Fiske & Taylor 2013; Spaulding 2017).

[170] In Chapter 7, I will go on to argue that models of belief also vary strikingly from belief attributor to belief attributor.

[171] On the most plausible reading he always was (Enger 2010; Freedman 1994; Pryal 2010): in *To*

In *Go Set a Watchman* (2015: 245–246), Harper Lee's recently published first pass at the setting, characters, and themes of *To Kill a Mockingbird*, Atticus makes his racist beliefs explicit.[172] Defending his leadership role in citizens' councils opposing the activism of the NAACP, Atticus asks Scout, "do you want Negroes by the carload in our schools and churches and theaters? Do you want them in our world? … Do you want your children going to a schools that's been dragged down to accommodate Negro children?" Atticus does not desire these things, because he believes that "white is white and black's black." "So far," he says, "I've not yet heard an argument that has convinced me otherwise." He chastises Scout for not sharing this belief; "Honey, you do not seem to understand that the Negroes down here are still in their childhood as a people. You should know it, you've seen it all your life."

Atticus—indisputably, if *Watchman* is canonical—believes that black people are intellectually inferior to white people. But here is the thing. He does not believe it in the same brute style as Robert E. Lee Ewell, the villain of *Mockingbird*. Whereas Bob Ewell's racism is crass, dehumanizing, and hateful, Atticus's racism is studied, condescending, and paternalistic. While both men sincerely and explicitly believe in the intellectual

---

Kill a Mockingbird, Atticus reassures his daughter, Scout, that he's "about as radical as Cotton Tom Heflin" (1960: 336), the white supremacist senator from Alabama.

[172] In treating Atticus's (and Ewell's and Juliet's) racism as a matter of belief, I am following Tommie Shelbie (2016), who defines "racism [a]s an ideology: a widely held set of associated beliefs" (22), and writes that "someone who explicitly subscribes to a racist belief system is obviously a racist, but so is someone who is disposed to act on racist assumptions (even when the person does not fully know that such assumptions shape his or her conduct and attitudes). A racist action is one undertaken because of the agent's racist attitudes or an action the agent rationalizes in terms of racist beliefs" (24).

inferiority of black people, they live out this belief in dramatically different styles.

Case #2: Soylent associated

People have different styles of believing much simpler and more specific beliefs as well. Consider Dennett's various ways of believing that a Frenchman has committed murder in Trafalgar Square. Or consider this: I put a bottle of Soylent in the Penn Philosophy Department refrigerator before writing this paragraph. I have presumably just led my (countless) readers to form the same belief, and even to form that belief in the same way. Everyone reading this passage now believes that there was Soylent in the fridge at the time of this writing. But readers may also have a wide range of levels of experience with, knowledge of, and interest in Soylent.

Imagine that a reader named Marie had never heard of Soylent before. Because she trusts me, Marie does now believe, of Soylent, that there has been some in the department fridge. She has inferred some things about Soylent from the facts that is refrigerateable, comes in a bottle, and is called 'Soylent'. But that is all she has to work with. Now imagine that another reader, Patrick, is wild about Soylent. He prefers its pancake battery taste to that of any other drink. He feels ten years younger since he replaced his diet with it wholesale. When somebody mentions that there might be even more Soylent to be acquired, his ears perk up. He has memorized the ingredient list of Soylent 2.0—the kind that, as he knows, comes in a bottle—and can tell you exactly how it differs in make-up from Soylent Powder, Soylent Coffiest, and the Soylent Bar. He

even has fond memories of making friends with the UPS delivery person who makes sure his Soylent subscription arrives safe and sound every two weeks.

Patrick and Marie both believe that there is Soylent in the fridge. We can veridically attribute that belief to both of them. But Patrick's belief is more interesting—associated with a much richer range of dispositions—than Marie's. Like Atticus and Ewell, Patrick and Marie have dramatically different styles of believing the same thing.

3.2. Styles of belief

Now, Juliet also believes that black people are intellectually inferior to white people, insofar as she does believe it, in a dramatically different fashion than Bob Ewell. Unlike Juliet, however, Atticus Finch would be poorly described as merely in-between believing that black people are intellectually inferior to white people. Atticus's bias is not an implicit bias. He severally explicitly avows the belief that black people are intellectually inferior to white people, and his actions with the Maycomb Citizens' Councils speak even louder than his words. (The same holds true in the mundane Soylent case. Marie and Patrick both explicitly—and not just sort of—believe there is Soylent in the fridge.)

Even so, Atticus and Ewell are disposed to behave, think, and feel very differently from each other, in accordance with their shared belief. Ewell is disposed to mutter (and sometimes scream) racial slurs whenever a black person does something he thinks is stupid. Atticus does not swear in public, and never uses nasty slurs. He even,

famously, has some of the dispositions stereotypically associated with being anti-racist: he cherishes and trusts his housekeeper Calpurnia, he works hard to defend his client Tom Robinson, he raises his children to treat people of all races with respect, and so forth. Nevertheless, he is disposed to engage in political battles against integration in Alabama schools. And, again, if you just sit down and ask him, he will sincerely tell you he thinks black people are intellectually inferior to white people.

Some (though far from all) of the difference in style of belief between Atticus and Ewell can be chalked up to the fact that they believe for different reasons. Atticus believes that black people are intellectually inferior to white people at least in part because he thinks that they are culturally inferior. Ewell, on the other hand, thinks that black people are naturally inferior to white people. These differences in reasons yield further differences in dispositions. For example, Atticus might be disposed to allow that black culture in the United States could mature to the point where it would be reasonable to integrate American schools. Ewell, on the other hand, is disposed to vehemently deny this possibility, and assert that God made black people naturally inferior to white people, like nonhuman animals.

Marie, meanwhile, recently became disposed to answer 'yep' when asked if there was Soylent in the fridge, to be able to name the beverage in a white bottle when she sees it in the fridge, and so, boringly, on. Patrick gained a much richer set of dispositions, though he obtained the same belief as Marie and believes it for the same reason (they both accepted my testimony). Patrick became disposed to salivate when he

approaches the fridge, pay more attention to my arguments for Ryleanism, answer '2.5' when asked how many grams of polyunsaturated fat I put in the fridge earlier, and vividly imagine sharing yet another cold, freshly refrigerated bottle with his UPS delivery person.

The question is: how is it that people who explicitly believe the very same proposition—that black people are intellectually inferior to white people, or that there's Soylent in the fridge—can be disposed to think, feel, act, and react in accordance with that shared belief, but in such strikingly different manners?

3.3. Styles of believer

Paramechanical functionalists about belief have a relatively straightforward way of answering this question. Functionalists tend to specify the functional roles that individuate beliefs in terms of a disjunction of counterfactuals, varying with respect to the other possible mental states of believers. Thus, functionalists can take different believers' respective beliefs that *p* to be individuated in exactly the same (disjunctive, counterfactual) way—that is, to play the same (disjunctive, counterfactual) functional role in the believers' respective cognitive systems—and account for the differences between the two men in terms of their other attitudes.[173] On this paramechanical

[173] Of course, paramechanical functionalists can only get away with this move if they can vindicate paramechanical functionalism in the face of the dissociation that styles of belief impose between attitudes of belief and cognitive states of belief. I argue for that dissociation in Chapter 4, Sections 4.3.2 and 4.3.3.

functionalist view, Atticus and Ewell do share the same belief in the intellectual inferiority of black people; this belief just interacts with their other, divergent, beliefs, desires, values, and memories in order to produce divergent behaviors. Similarly, the paramechanical functionalist would diagnose Marie and Patrick as having the same belief that there is Soylent in the fridge; they just differ in their other attitudes (for example, only Patrick also believes there is Soylent 2.0. in the fridge and desires a taste thereof). According to paramechanical functionalism, then, believers do not believe that *p* differently; they merely think, feel, and behave as if *p* differently on account of how their shared belief interacts with other components of their mental profiles.

Schwitzgebel does not have recourse to this kind of explanation of these cases, since Ryleans reject the assumption of the productive causal powers of attitudes of belief implicit in the paramechanical functionalist account. But Schwitzgebel could, in principle, take a lesson from the functionalists by building other mental attitudes into the activation conditions for the dispositions that divide Marie and Patrick. Indeed, some attitudes must be so built in. For example, on any Rylean account, many beliefs that *p* are partly constituted by the disposition to sincerely assert that p, and the activation conditions for that disposition include, among other things, the desire (or at least willingness) to let it be known that *p*.

Nevertheless, this pseudo-functionalist approach could only take Schwitzgebel so far, for two reasons. First, there is no plausible sense in which Atticus is disposed to use racial slurs. He seems neither to possess—nor to be prone to come into possession

of—any of the values and desires which would serve to activate such a hypothetical disposition. To say that Atticus is so disposed, but never triggers the activation conditions, would be to invite a slippery slope. Every human on earth could be said to be disposed to do everything within their capacities—if not everything *humanly* possible—if the activation conditions for dispositions were construed broadly enough.

Second, as a good Rylean, Schwitzgebel (2013) also gives a dispositional analysis of *desires* to let it be known that *p*. As it turns out, desires to let it be known that *p* will usually be determined with respect to a dispositional stereotype that includes dispositions that also partly constitute beliefs that *p*. Schwitzgebel rightly explains that

> one remarkable feature of stereotypes is that they overlap. The same dispositions can belong to more than one stereotype. Consequently, by virtue of satisfying one stereotype, a person can, for free, nearly match the dispositional stereotype for a closely related attitude. (2013: 21)

For example, Atticus's belief that black people are intellectually inferior to white people is partly constituted by many of the dispositions that also constitute his desire for segregated schools. This overlapping of dispositional stereotypes means that the differences between believers cannot be spelled out entirely in terms of differences in their other attitudes. On a Rylean understanding of the attitudes—which must be holistic if it is to avoid collapse—differences in believers' other attitudes are very often

ipso facto differences in the dispositions that make up their beliefs themselves.

Both of these obstacles preventing Schwitzgebel from explaining styles of belief solely in terms of other attitudes derive their force from within a Rylean understanding of attitudes of belief (and the dispositions that they comprise).[174] In particular, both reasons hinge on the Rylean insight that the character of believers is 'molar', rather than 'molecular'.[175] Character would be molecular if its components—habits, traits, attitudes, skills, memories, intelligence, imagination, and so on—could be understood independently of each other. However, as Ryle stressed, there is no clean way of carving up character. Elizabeth Bennett and Mr. Darcy are both prideful, but this does not mean they share an identical, molecular character trait of pride. On the contrary, their respective ways of being proud are thoroughly entrenched in their respective characters. Character is molar insofar as any given component of character can be fully understood only in light of the whole. Darcy's pride is properly labeled as 'belonging to Mr. Darcy' not just because the proprietor of Pemberley happens to be proud, but also because his style of pride inheres in his unique character. As Ryle writes,

> Jane Austen's people [like real people] are, nearly always, alive all over, all

---

[174] To be clear, I do not mean to insinuate that Schwitzgebel would take the pseudo-functionalist approach; he is clear that differences in belief are to be chalked up to differential molar fits to dispositional stereotypes.

[175] I have lifted and adapted this terminology from mid-20th century psychology. Gestalt psychologists (Koffka 1935; Köhler 1947) described experience as molar, and purposive behaviorists (most prominently Tolman 1932) described behavior as molar.

> through and all round, displaying admirably or amusingly or deplorably proportioned mixtures of all the colours that there are, save pure White and pure Black. If a Calvinist critic were to ask us whether Mr Collins was Hell-bound or Heaven-bent, we could not answer. The question does not apply. Mr Collins belongs to neither pole; he belongs to a very particular parish in the English Midlands. He is a stupid, complacent and inflated ass, but a Sinner? No. A Saint? No. He is just a ridiculous figure, that is, a figure for which the Calvinist ethical psychology does not cater. (285)

Ryle's point is that Collins's particular attributes—his stupidity, complacency, and so on—cannot be fully understood except in light of his entire (ridiculous) character. The same goes for beliefs. Atticus's belief that black people are intellectually inferior to white people—his particular style of believing this proposition—is fully explicable only qua *Atticus*'s belief, with all of Atticus's idiosyncrasy of character entailed. Correspondingly, beliefs are rarely our ultimate objects of inquiry. People usually attribute beliefs to themselves and one another in order to better understand (or manipulate, praise, or sanction…) whole believers. Readers of *To Kill a Mockingbird* and *Go Set a Watchman* are interested in the details of Atticus's racist belief, for instance, because it helps us understand Atticus himself.[176]

---

[176] This molar conception of character is one reason that I am inclined to attribute intentional content to whole believers, rather than particular beliefs, as discussed in Chapter 5.2.7. If character were molecular, beliefs might have discrete contents. Since character is molar, it makes

This molar conception of character provides the Rylean with a clear heuristic (although not anything like a precise metric) for distinguishing the lack of a disposition from the existence of a disposition with chronically unfulfilled activation conditions. Believers have dispositions with chronically unfulfilled activation conditions insofar as they are the type of people who might fulfil—that is, insofar as they possess characters conducive to the fulfillment of—those activation conditions. Believers lack dispositions insofar as they are not that type of person: insofar as fulfilling the activation conditions would involve an attendant change in (or at least lapse of) character. Again, the preceding is not a precise metric by which to decide whether somebody has a perpetually latent disposition or lacks that disposition. However, it is an intuitive heuristic by which we can veridically judge clear cases. Patrick has the disposition, which Marie lacks, to imagine sharing a cold Soylent 2.0. with a good friend, because he has the relevant memories and knowledge (and capacity for vivid imagination). If Marie were to gain those memories and knowledge, her personality profile would be relevantly altered, and she would gain that disposition too. Atticus, meanwhile, lacks the disposition to use nasty slurs because he is not that kind of man. To fulfill the activation conditions for slurring would be to become a different kind of man. On the other hand, even though Atticus Finch has never reprimanded his daughter for dating a black man, *Watchman* provides solid evidence that he would do so if the activation

---

more sense to consider belief attributions to alter attributors' molar conceptions of how whole believers take the world to be.

conditions arose. Such an action would be (unfortunately but undeniably) in character. Thus, Atticus has the disposition to so act.

Divergent dispositions are the stuff, and not the mere trappings, of our protagonists' respective beliefs. As Ryle writes,

> … styles and procedures of people's activities *are* the way their minds work and are not merely imperfect reflections of the postulated secret processes which were supposed to be the workings of their minds … Overt intelligent performances are not clues to the workings of minds; they are those workings. Boswell described Johnson's mind when he described how he wrote, talked, ate, fidgeted and fumed. (1949: 58)

It is not just the results of belief that must be analyzed relative to entire characters; it is the constitution of believing itself. From this perspective, it does no good to construe the tendency to use nasty racial slurs (given the appropriate activation conditions) as a perpetually unactivated component of Atticus's belief. His belief is a modification of his character, and his character is diametrically opposed to fulfilling the activation conditions for using nasty racial slurs. His character is nevertheless perfectly consonant with the (as yet unfulfilled) activation conditions for scolding his daughter for her romantic interest in a black man. The latter disposition is the stuff of his belief that black people are inferior to white people; the former is not.

Thus, while Atticus differs from Ewell in some of his other attitudes, he also—by the very same token—differs from Ewell in his style of belief. Schwitzgebel has detailed one way in which believers can differ in their style of belief: they can differ in the degree to which they believe (not just in degree of confidence—as Bayesians would have it—but also in degree of fit to a dispositional stereotype). However, unlike Juliet, Atticus does not believe that black people are inferior to white people to any lesser degree than Ewell.

Elizabeth Bennett is no less proud than Darcy, but she does have a profoundly different style of pride. If pride, considered in general, is the trait of living as if one deserves respect, then Ms. Bennett and Mr. Darcy vary in the styles in which they are disposed to think, feel, and behave as if they deserve respect. Analogously, Atticus and Ewell vary in the styles in which they are disposed to think, feel, and behave as if black people are inferior to white people. Part of this difference in style of belief amounts to a difference in the other attitudes the two men possess. Nevertheless, for Ryleans who embrace the molar conception of believers as whole persons, this difference in other attitudes must also amount, equally, to a difference in the constitution of belief.

3.4. A (Ryle-on-Austen-ian) revision of Schwitzgebel's account

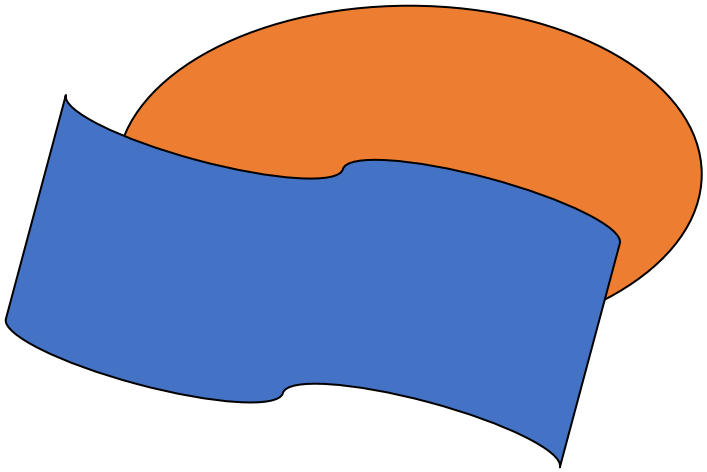
Schwitzgebel's conceptual machinery does not provide a satisfactory answer to the question of how to account for diversity in style of belief. Schwitzgebel posits a single general-purpose stereotype for each belief, against which every potential believer is measured. This stereotype does not comprise an exhaustive list of all of the ways in

which believers might act (or react) as if *p*. Instead, it comprises only those dispositions that a normal interpreter would be apt to associate with the belief that *p*. Every believer fits this stereotype more or less well. Thus, Schwitzgebel has a good answer to the question of how Marie and Patrick can be said to share the belief that there is Soylent in the fridge: they both fit the general-purpose stereotype well enough to count as having that belief. But Schwitzgebel does not have a good answer to the question of how it is that they live out their respective beliefs in such dramatically different ways.

It might be thought that Schwitzgebel covers this base by pointing out that believers diverge from dispositional stereotypes in certain *respects*, as well as to certain degrees. But vague talk of respects of divergence does not do the trick. For starters, it seems implausible that several of the dispositions that make up Patrick's particular style of believing are dispositions that a normal interpreter would be apt to associate with the belief that there is Soylent in the fridge. Patrick's disposition to imagine hanging out with his UPS delivery person is part and parcel of his belief that there is Soylent in the fridge: if you know Patrick well, then you know that whenever he believes Soylent is around, he starts daydreaming about his buddy Greg. However, this idiosyncratic disposition of Patrick's will (rightfully) be nowhere to be found in most Schwitzgebelian general-purpose stereotypes of the belief that there is Soylent in the fridge.

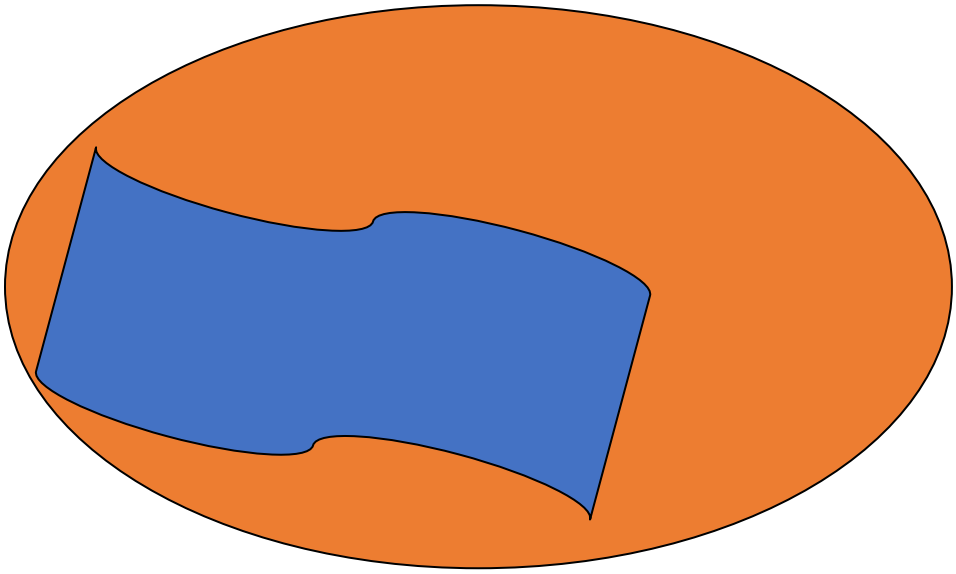
More abstractly, Schwitzgebel's general-purpose dispositional stereotypes lack the proper metaphorical s*hape* to shed light on the particular styles in which different people believe that *p*. Consider Figure 10.

FIGURE 10: A GENERAL-PURPOSE STEREOTYPE AND A STYLE OF BELIEF



If the orange shape in Figure 10 represents the general-purpose stereotype for the belief that there is Soylent in the fridge, and the blue shape represents Patrick's dispositions to think, feel, act, and react as if there is Soylent in the fridge, then it does not seem best to describe Patrick's belief just by saying how he conforms to and deviates from the general-purpose stereotype. In line with my point in the previous paragraph, the blue shape exceeds the boundaries of the orange shape in crucial respects. The blue shape also hangs together in its own unique way. Patrick's belief is not best defined by its relationship to the orange shape; it has its own structural integrity. There are many things to say about how Patrick believes that there is Soylent in the fridge that are not well captured by Schwitzgebel's vague language of fit to the general-purpose stereotype for that belief. This point holds even if the general-purpose stereotype in question is construed extremely broadly, as in Figure 11.

FIGURE 11: A VERY BROAD GENERAL-PURPOSE STEREOTYPE AND A STYLE OF BELIEF



Even if, against my expectations, the normal observer is apt to associate all sorts of weird but possibly relevant dispositions with the belief that there is Soylent in the fridge, a measure of how much and in which respects Patrick deviates from the full stereotype is not the best—or, importantly, the typical—way for attributors to come to understand Patrick's unique style of believing. Why, as lay belief attributors, would we ever construct such a fully general model of belief? And even if we did, why would we use it in order to pin down a style of believing?

By the same token, Atticus's belief that black people are intellectually inferior to white people *can* be analyzed with regard to conformity and deviation from the general-purpose dispositional stereotype for that belief; however, such an analysis is not the most informative way to come to understand the idiosyncrasies of his belief. Different

dispositions are more central, and different dispositions more peripheral, to Atticus's belief than to the stereotypical racist belief. Moreover, unlike Juliet, Atticus does not somehow fall short of full belief insofar as he deviates from the general-purpose stereotype. The fact that he is in no way prone to use nasty racial slurs, for instance, does not reliably indicate that he is any closer to believing in the intellectual equality of the races than if he were prone to using such slurs.

General-purpose stereotypes are most effectively employed as initial tools for sizing up and categorizing believers. To achieve a serious practical understanding of any particular person as a believer, belief attributors must move beyond general-purpose stereotypes and, to paraphrase Ryle on Austen, pin-point the exact quality of the belief, and the exact degree of that quality, by matching it against the same quality in different degrees, against simulations of that quality, against deficiencies of it and against qualities which, though different, are brothers or cousins of that selected quality. According to Schwitzgebel's machinery, the most informative answer to 'does Atticus believe that black people are inferior to white people?' is 'sort of', or perhaps 'yes, though he deviates from the stereotype to this degree and in these respects'. I disagree with Schwitzgebel that these are the most informative answers. 'Sort of' is just a stop-gap answer: an answer we give while still working with the blunt tool that is a general-purpose stereotype.

Schwitzgebel acknowledges that when we really want to understand somebody, we move beyond dispositional stereotypes and develop a description of the person's

own idiosyncratic pattern of dispositions. However, Schwitzgebel claims that in such scenarios "belief language starts to break down; the simplifications and assumptions inherent in it aren't entirely met" (2008: 535). When belief attribution via a general-purpose stereotype falls short of providing us with a fully satisfactory understanding of the believer, we must move past the attribution of belief—which, on Schwitzgebel's view, is just "shorthand" (2002: 271) for the direct attribution of dispositions anyway—and try to get a direct grip on the messy dispositional details. Schwitzgebel writes that "once all the relevant dispositions have been made clear, the case is closed. There are no further facts to report" (2002: 262).

Belief attribution is much more powerful—and lay attributors' models of belief much less rigid—than Schwitzgebel's conceptual machinery allows. By discussing only general-purpose stereotypes, on the one hand, and unorganized morasses of dispositions on the other, Schwitzgebel both fails to draw attention to the impressive flexibility inherent in human belief attribution practices and fails to capitalize on the fact that patterns of dispositions have interesting emergent properties over and above the properties of mere unstructured lists of dispositions. Contra Schwitzgebel, even if all of somebody's relevant behavioral, cognitive, and phenomenal dispositions have been made clear, the case is far from closed; the further facts to report are the attitudes of belief themselves: the emergent and irreducible *patterns* of dispositions that are believers' more or less idiosyncratic styles of living as if *p*.

Models of belief do serve to capture swaths of individual dispositions at once,

but their most important purpose is to outer-recognize real patterns that figure in molar characterizations of whole believers. As discussed in Chapter 3.3.2., model theorists understand that in order to achieve this aim, belief attributors must construct and refine two broad varieties of model of belief: in addition to general-purpose stereotypes—which provide attributors with "an understanding of a general structure or schematic pattern that can have many specific instantiations" (Godfrey-Smith 2005: 4)—attributors construct models of how particular believers (in particular contexts) believe. These specialized models—some of which "are extremely fragmented and minimal, while others are rich and detailed" (6)—allow attributors to understand individual believers on a more nuanced level. Both varieties of model outer-recognize real patterns. General-purpose stereotypes outer-recognize relatively generic patterns that are lived out in diverse manners by many believers; specialized models outer recognize relatively stylized patterns that are lived out in more specific manners by fewer believers. The former serve mainly to reveal commonalities between believers that *p*; the latter serve mainly to highlight styles of belief. Lay belief attributors routinely apply both types of model to attribute beliefs to their friends and acquaintances.

Taking a note from Heidi Maibom (2003, 2009) and Peter Godfrey-Smith, a better answer than 'sort of' to the question 'does Atticus believe that black people are inferior to white people?' is 'yes; he believes that black people are inferior to white people in this particular style', followed by a qualitative description of the pattern in which Atticus Finch lives by that belief. This answer requires the interpreter to go beyond simply

measuring Atticus up against a standard, one-size-fits-all general-purpose stereotype for believing that black people are inferior to white people. However, it does not require the interpreter to cast aside the notion of belief altogether, and delve all the way down to the dispositional details. Instead, it requires the construction and consultation of a brand new, believer-specific dispositional stereotype: a model of believing that black people are intellectually inferior to white people *in the style of Atticus Finch*.

This new believer-specific model will differ from the general-purpose stereotype in (at least) three respects. First, it will include dispositions that the general-purpose stereotype does not; for example, Atticus's paternalistic style of believing partially comprises the disposition to be extremely polite towards black people. Second, it will omit dispositions that the general-purpose stereotype includes; Atticus is not prone to using nasty racial slurs. Third, it will mark different dispositions as central to the belief; the tendency to worry about the political and legal system being corrupted by the influence of black voters may be peripherally associated with the general-purpose stereotype, but it is absolutely central to any good believer-specific model of Atticus's style of belief.

3.5. General-purpose stereotypes and believer-specific models

We construct and consult believer-specific models and attribute belief in this fine-grained manner all the time. LeBron James believes he is the greatest basketball player to ever live, but he does not believe it like Michael Jordan. Nobody believes in

themselves quite like Mike. Indeed, both Jordan's success on the world stage and his failure at more intimate human endeavors are often chalked up to his particular, monomaniacal and vainglorious style of believing in his own greatness (Lazenby 2015). A single mother from West Philly, a Syrian refugee, and Donald Trump may well all believe–and not just kind of believe–that life in America is hard, but belief attributors easily recognize that their respective beliefs comprise dramatically different patterns of dispositions. If we really want to understand what these individuals are like, via attributions of belief, then we have to get much more specific. Returning, then, to our central case: to understand that Bob Ewell believes (that black people are intellectually inferior to white people) in a crass, dehumanizing, and hateful manner, whereas Atticus Finch believes it in a studied, condescending, and paternalistic manner, is to begin to understand the two characters for who they really are.

Proliferating models of belief allows belief attributors to account smoothly for how Atticus's belief is both the same as—and radically different from—Ewell's. Consider the analogy between believers and metamers invoked in Chapter 4. Like metamers, Atticus and Ewell both believe the same thing, when seen in one metaphorical light. Given their divergent reflectance properties, metamers manifest the same perceivable color in different ways. Likewise, given their divergent molar characters, Atticus and Ewell have different styles of holding their shared racist belief. When the metaphorical lighting changes, and we start paying attention to what is unique about them rather than what they share, Atticus and Ewell can be seen to believe

in dramatically different ways.[177] The real patterns in which the two men express their attitudes towards race both overlap substantially and come apart in significant respects. Whether belief attributors should focus on the overlap or the divergence—and thus work with general-purpose or believer-specific models—depends largely on the context in which attributors are concerned to figure out what they believe.

Merely to attribute a belief that *p* to somebody is to explain very little about them. General-purpose stereotypes are far from without their uses: sometimes people do not need to know very much about somebody in order to predict, explain, judge, or regulate their behavior, and sometimes human beings are simply obsessed with distinguishing the believers from the non-believers. Indeed, people frequently rely on general-purpose stereotypes even though believer-specific models are ready to hand. Aristotle, Averroes, Aquinas, Astell, and Anscombe all believe in God. That is not to say much about their respective religious views, but it is the right attribution when the attributor's sole purpose is to divide the atheists from the believers.

Attributors also construct and consult specialized models at various levels of generality between general-purpose stereotypes and believer-specific models. Just as

---

[177] Or consider a different analogy between belief and color. Red things vary in shade. All red things are determinably red, but what determinate shade each red object has varies hugely. Red things are all the same color only according to a broad criterion of what it takes to be the same color, though we often—and felicitously—speak of color in terms of this broad criterion. Similarly, believers that *p* share the determinable belief that *p*, but what determinate belief each believer possesses varies hugely. Atticus Finch and Bob Ewell's are the same belief only according to a general-purpose model, though we often—and felicitously—speak of belief in terms of this general-purpose model (for the purposes of everyday life and the social and human sciences).

perceivers can group one subset of shades of red as crimson, and another subset as scarlet, belief attributors can (and probably should) note that Ewell and Atticus both believe that black people are inferior to white people in the style of white US southeasterners in the middle of the 20th century. Similarly, a few of the varieties of belief distinguished in Chapter 2 arise in relation to specialized models of belief. For example, some specialized models are models of attitudes of assent (de Sousa 1971; Dennett 1978, 1991; Cohen 1992; Frankish 2004): patterns of dispositions that attributors identify with taking linguistically structured propositions to be true. In many analytic epistemology seminars, philosophers cast aside general-purpose models of belief in favor of working solely with specialized models of assent.

We also sometimes make generalized use of believer-specific models. Patricia Churchland, Daniel Dennett, Jerry Fodor, and Gary Hatfield all believe that empirical results are relevant to philosophical inquiry, but they believe it in different styles. (I wield believer-specific stereotypes for each of their particular beliefs.) David Chalmers, meanwhile, believes that empirical results are relevant like Fodor (and in some ways like Hatfield, and a bit like Dennett), but not so much like Churchland. That is to say: he fits Fodor, Hatfield, and Dennett's stereotypes better than he fits Churchland's, though of course there is a believing-that-empirical-results-are-relevant-to-philosophy-like-David-Chalmers stereotype to be constructed that fits him best of all.[178]

---

[178] As argued in Chapter 4, this proliferation of models of belief drives a wedge between attitudes of belief and cognitive states of belief, even when cognitive states are understood as person-level functional states. The contours of attitudes of belief are determined (not by the functional roles

In sum, I submit that there are (at least) as many potential models of the belief that *p* as there are believers—and classes of believers—that *p*. What everybody who believes that *p* has in common is that they all cross the threshold for believing that *p*, according to a general-purpose dispositional stereotype for that belief. But describing somebody merely as believing that *p* is not to describe their attitude of belief that *p* in full (or even much) detail. To describe any given believer's attitude of belief in full detail would be to say a great deal about that believer that they do not (necessarily) have in common with any other believer that *p*. However, contra Schwitzgebel, such a description would still pick out an emergent pattern, rather than a mere tabulation of individual dispositions.

4. Lack of character?

Over the last 25 years, philosophers (Flanagan 1991; Harman 1999, 2000; Doris 2002; Alfano 2013) have impugned dispositionalist accounts of character in order to challenge the (purported) moral psychological basis of virtue ethics. The so-called situationist challenge to virtue ethics is straightforward. Decades of experimental social psychology have revealed that how people behave depends, to a large degree, on (apparently irrelevant) situation factors. People who smell cookies give more to charity (Isen & Levin 1972); seminarians who are running late are less likely to help out

---

they play in cognitive systems but) by the way they fit *particular* models of belief, just as the perceived colors of metamers are determined by how the light they reflect is perceived by particular perceptual systems.

homeless people—even when they are late for a lecture on the Good Samaritan (Darley & Batson 1973). The situationist challenge weaponizes these empirical findings. Since how people behave depends so much on the situations in which we find ourselves, it must not depend very much—certainly not as much as virtue ethicists would like to think—on our virtues and other character traits. Indeed, since character traits are typically analyzed dispositionally—in terms of how we regularly behave—we must not really have virtues at all.[179]

This challenge might naturally be extended to my dispositionalist account of belief.[180] How believers behave is (1) overwhelmingly situation-dependent, and (2)

---

[179] Virtues—in order to count as virtues—are particularly stable and consistent (as well as laudable) character traits (Alfano 2013: 26). The situationist challenge is usually framed as a challenge to the idea that people are virtuous; but to be extended to my dispositional account of belief, or to properly challenge dispositional accounts of character traits in general, as opposed to virtuous character traits alone, it has to be slightly reframed. The challenge must be that behavior's situation-dependence undermines it being explained by reference to character. But (to foreshadow my argument in the main text) it does not seem plausible that all—or even most—character traits are cross-situationally consistent in the manner of the virtues. Consider some vices. Socrates, being virtuously intellectually modest, may never claim to know something he does not know. But somebody who is intellectually arrogant does not always claim to know things she does not know. She is simply disposed to pretend to knowledge in certain situations—particularly situations in which she thinks she can get away with it. Similarly, the skilled inveterate liar does not lie about everything (even though the honest man may always tell the truth). Whether or not she is disposed to lie is extremely situation-dependent. And it is not only vices that tend to lack consistency—but still qualify as character traits—in this manner. Consider the introvert (or extravert, or, for that matter, bearers of any of the other Big Five personality traits). Alternatively, consider the chatty person, who might be an introvert or an extravert, but likes to engage in small talk in social situations she finds comfortable. (Note, too, that there are vicious and virtuous styles of being introverted, or chatty.)

[180] The situationist challenge is a potential problem for virtue ethicists who adopt Schwitzgebel's original conceptual machinery, but its potential grows given the arguments of this chapter: Atticus's particular style of belief is even less stable than his adherence (or not) to the general-purpose stereotype. Styles of belief are shifting all the time; whether or not a believer reaches the threshold for believing that p is more stable, since shifts in the dispositions that make up a belief may make no difference to whether the belief remains well above the threshold for believing.

varies widely in its situation-dependence from subject to subject (and corresponding style of belief to style of belief). Styles of belief are also (3) constantly (if subtly) in flux. Moreover, people (4) often fail to live up to the beliefs that are reasonably attributed to them, including the beliefs they reasonably attribute to themselves. These four facts do provide a challenge to any virtue ethics built around a Rylean framework. Nevertheless, far from being a challenge to Ryleanisms that embrace styles of belief themselves, these same four facts reveal the importance of the notion of styles of belief for the Rylean project. If, in so doing, they undermine the hope for a neat, context-independent, and fully general virtue ethics, then so much the worse for such a hypostasized normative theory. Ryleanism about belief walks away at worst unscathed, and at best significantly richer. Consider the four purportedly challenging facts in turn.

First, whether or not somebody acts in accordance with a belief is paradigmatically situation-dependent. On my view, Rylean accounts of belief should embrace this situation-dependence. The situationist John Doris has proposed that virtue ethicists refocus their efforts on a notion strikingly similar to styles of belief: 'local character traits', such as fair-weather-generosity, as opposed to generosity in general. Doris argues that people possess only local traits.[181] However, while Doris is right that more attention should be paid to local, fine-grained traits, he is wrong that local traits can serve to completely replace cross-situational traits in moral psychology. Fair-

[181] See also Adams 2006 and Miller 2013; but see Harman 1999 and 2000 for arguments that local traits should not be considered 'traits' at all.

weather-generosity is a style of generosity; whether the fair-weather-giver counts as all-things-considered-generous depends on whether or not she sufficiently resembles the general-purpose stereotype for generosity. It might well be that many virtues, including generosity, entail a rare consistency of behavior across situations. Moreover, it might well be that many people who are fair-weather generous (per the general-purpose stereotype for fair-weather-generosity) fail to be generous *tout court* (per the general-purpose stereotype for all-things-considered generosity). This general failure to fit the stereotype does not show that we should do without general-purpose models of character traits. Lay people often employ general-purpose models in order to understand, manipulate, and categorize each other. The information that most people fail to act, react, think, and feel in sufficiently cross-culturally stable manners—and thus fail to sufficiently fit general-purpose models—is itself profoundly informative about the molar character of our friends and acquaintances.

Second, Mark Alfano criticizes Doris's notion of local traits because it

> suggests that different distinctions may have to be made for different people. In the same way that one wants to be able to say (and learn) that *A* and *B* have the same belief, desire, thought, or emotion, one wants to be able to say (and learn) that they have the same virtue and vice. If, however, local traits are individuated in such an ad hoc, fine-grained way, this may be impossible. The local trait theorist would have to add that the fine distinctions made in one person's case

> would for the most part apply in another person's case. Else, the traits that *A* has might turn out to be incommensurable with the traits that *B*, *C*, and *D* have. (2013: 66–67)

Contra Alfano, fine-grained distinctions in people's local traits (or styles of belief) do not mean that we cannot say (or learn) that both people nevertheless possess—or are somewhere in-between possessing and not possessing—the general, coarse-grained trait (or belief). Paramechanical functionalists may have a hard time saying that *A* and *B* have the same belief, if beliefs are individuated in an ad hoc, fine-grained way. But Ryleans, with recourse to the distinction between general-purpose stereotypes and believer-specific models, should have no such difficulty. *A* and *B* share the same belief relative to the general-purpose stereotype, but their individual beliefs are individuated relative to believer-specific models. The fact that different people can believe the same thing, despite being systematically disposed to express that belief in different situation, coheres perfectly with the notion of styles of belief developed in this chapter.

Third, styles of belief can change in subtle ways. People do not only change with regard to whether or not they believe something; they also frequently change the manner in which they believe it. Once again, my Ryleanism provides the resources to account both for changes in the dispositions that make up somebody's unique (and uniquely evolving) style of belief, and for the fact that those changing dispositions could make up a diachronically stable belief that *p* over time. In order to fully understand

what somebody believes, belief attributors make frequent (if subtle) changes to their believer-specific models (in response to the believer's shifting pattern of dispositions), while simultaneously recognizing that the believer has continuously appropriately matched the general-purpose stereotype for believing that *p*.

Fourth, belief attributions (like character trait attributions) are partly aspirational (Tumulty 2014; Callard 2018), or "morally factitious" (Alfano 2013). That is: interpreters often attribute the traits and beliefs that they want people to live up to, rather than the traits and beliefs that they take people to actually possess. Nevertheless, there is no evidence that bravery and pride are not—or, to be more precise, cannot be—real. It is just extremely difficult for people to be consistently brave or prideful across contexts. Similarly, it is not the case that particular cross-culturally stable styles of believing do not—or cannot—really exist. It is simply difficult for people to believe consistently across contexts. Often, when we attribute belief, we take believers to approximate the ideally consistent believer with greater fidelity than they really live up to. Our attributions of belief in these cases *are* often (more or less) nonveridical. What an aspirational believer really believes—their true style of belief—is messier, less consistent, and more context-dependent than aspirational belief attributors give them credit for. Nevertheless, there is some attributor-relative fact of the matter about what they believe—some messy, inconsistent way in which they match belief attributors' aspirational models—even when it better suits belief attributors' practical purposes to eschew veridicality in favor of attributing an idealized style of belief.

If successful, the situationist challenge impugns a strong conception of saint-like virtue or thoroughness of belief; that is, it shows that most believers are only in-between believers, flirting with the threshold for meeting a general-purpose stereotype. Even so, the situationist challenge represents no threat to Ryleanism about belief. Instead, it contributes to the Rylean's positive empirical understanding of how situation factors (as well as personal factors) influence traits and styles of belief. Where Doris sees fragmentation of character and delusional interpretations, I see rich cognitive differences and commendably aspirational (if sometimes resultantly nonveridical) belief attribution practices.[182]

People are complicated. We resemble and differ from each other in myriad subtle ways. Moreover, as Zadie Smith remarked in the wake of the Trump election,

> if novelists know anything it's that individual citizens are internally plural: they have within them the full range of behavioral possibilities. They are like complex musical scores from which certain melodies can be teased out and others ignored or suppressed, depending, at least in part, on who is doing the conducting (Smith 2016: 38)

Doris would nod approvingly at Smith's diagnosis. (It is in his intellectual character.)

---

[182] Cf. Doris's (2002: 62–91) discussion of the Social-Cognitive approach.

For, as Smith and Doris rightly remind us, how people are disposed to behave is overwhelmingly determined by their situations, including their social relationships. But this situation-dependence of behavior does not entail that people have incoherent or fragmented characters, much less that people lack character all together. Nor does it entail that we ought to explicate any given personality in terms of the full range of behavioral possibilities, considered independently of the situations in which those personalities happen to be embedded.

On the contrary, the situation-dependence of character means that individual personalities—including individual styles of belief—are usefully explicable only in terms of dispositions to act, react, think, and feel in *relevant* situations. (As I shall argue in Chapter 7, which situations are relevant depends in large part on *who* is asking, as well, perhaps, as *why* they are asking.) This situation-dependence is why Smith's novels—which, like Jane Austen's and Harper Lee's, centrally involve character studies—so often turn on placing big personalities in new contexts and watching how they respond. Atticus Finch's style of belief is only fully understandable in the context of a life embedded in Jim Crow Alabama. Atticus's belief does not include dispositions that might have been manifested by a different believer, living in a different place at a different time. It includes only those dispositions that constitute Atticus's tendency to live, in his place, during his time, as if black people were intellectually inferior to white people. To admit this situation-dependence is not to give up on the objective reality of belief. It is to embrace the extraordinary plurality that novelists recognize as latent in

human personalities.

5. Conclusion

To recap: by building on Schwitzgebel's notion of a dispositional stereotype, I have equipped Ryleanism with the tools to explain how it is that believers that *p* can vary dramatically in their styles of believing that *p*. Schwitzgebel argues that the biggest feather in Ryleanism's cap is that it usefully legitimizes in-between beliefs as objects of philosophical and psychological investigation, whereas paramechanical theories of belief do not. I would add that Ryleanism legitimizes general-purpose stereotypes, believer- and situation-specific models, and a whole range of more or less specialized models in between. Which models of belief are properly applied in any given context is determined not by a third-person functional analysis of the believer's cognitive system, but by particular belief attributor's second- or first-person purposes in attributing belief. Sometimes these purposes call for the application of general-purpose models that divide all of the believers from all of the nonbelievers. Othertimes, they call for a specialized model that more closely emphasizes an idiosyncratic style of belief. If any readers still do not believe that people exhibit many and various styles of believing, I have six good books to recommend them.

Chapter 7: Relativistic Ryleanism, intersubjectivity, and norms

1. Introduction

In Chapter 6, I advocated for a version of Ryleanism that accounts for different believers having different styles of belief. In this chapter, I am going to argue that particular believers can also have different beliefs in relation to different belief attributors. Consider the following parable.

1.1. Counseling the doubtful

The scene is the sacristy of a small Roman Catholic church in the southeastern United States. A parishioner has been having some doubts. The young priest and middle-aged deacon of the parish are discussing how best to counsel her. Over the course of their discussion, it becomes clear that they disagree about how to frame the parishioner's state of mind. They do not disagree about what she is thinking at any given moment, or how she is feeling. They both know how she behaves—not only how she moves her muscles, but the intentions with which she acts.[183] Nevertheless, they disagree about what she believes.

---

[183] For ease of exposition, I am stipulating that the priest and deacon agree about how the parishioner thinks, feel, and behaves on both thin descriptions—they agree about her tokens of mental speech, raw feels, and muscle movements—and thick descriptions—they agree about how she wholeheartedly genuflects, and skeptically arches an eyebrow while listening to certain gospel passages. As discussed in Chapter 5, this full intersubjective agreement about dispositions may be even harder to come by than it appears, since dispositions towards thickly described behaviors might be attributor-relative just like attitudes of belief.

The deacon emphasizes that "the faithful receive with docility the teachings and directives that their pastors give them" (*Catechism*: 87). The parishioner has not been receiving the teachings of her pastor with docility. On the contrary, she has been asking a lot of impertinent questions, especially on the subject of the resurrection. During bi-weekly bible study, she has stated that she is "not so sure about the whole Easter thing." The deacon takes her at her word and interprets her as disbelieving that Jesus rose again on the third day of his death. He thinks the goal of counsel should be to indoctrinate her: to get her to believe that Jesus was resurrected, from a place of disbelief.

The young priest interprets his parishioner differently. Yes, she has been asking a lot of questions. She has difficulties with the whole Easter thing. But "ten thousand difficulties do not make one doubt" (*Catechism* 157), and "faith seeks understanding" (158). More importantly, the parishioner attends mass every Sunday and professes her faith during the Apostle's Creed. (She also never misses a bible study or parish barbecue.) The priest therefore interprets the parishioner—a good practicing Catholic—as believing that Jesus rose from the dead.[184] After all, he takes thoughtful participation in the life of the Church to be the surest sign of belief. Even though she has some questions about how, exactly, Jesus could have risen again, the parishioner is "carried by the faith of others" in the parish to believe (166). So, the priest thinks the goal of counsel should be to strengthen the parishioner's extant belief in the resurrection.

---

[184] Boudry & Coyne (2016) defend the view that religious dogmas are beliefs (as opposed to imaginings).

The parishioner believes one thing from the deacon's point of view and another, contradictory thing from the priest's point of view. It is (at least) not obvious that one of them is mistaken.

1.2. Paramechanical denials of intersubjective indeterminacy

Paramechanists identify attitudes of belief with (computational, pure, or teleo)functional states (or their neural realizers) within the cognitive systems of believers.[185] Functional states are individuated with respect to their functions, and functions are not altered based on how they are interpreted from outside of the functional system.[186] Paramechanists thus deny that beliefs could vary between belief attributors. For example, they would deny that the parishioner could believe that Jesus rose again for the priest yet disbelieve it for the deacon. Instead, paramechanists would insist that there is a single fact of the matter about whether the parishioner has the belief that Jesus rose again.[187] The parishioner either believes, disbelieves, withholds belief, or

---

[185] Recall from Chapter 4.4 that there is room for teleofunctionalists to disagree about what the relevant functional system is: paramechanical teleofuctionalists individuate beliefs with reference to the adaptive histories of the cognitive systems of believers, whereas Rylean teleofunctionalists (like Dennett) individuate beliefs with reference to the adaptive histories of the social environments of belief attributors, which, qua functional systems, encompass both belief attributors' mindreading capacities and believers' mindreadable characteristics.

[186] Systemic capacity teleofunctionalists about cognitive states of belief—who take cognitive states to emerge within analyses of cognitive systems that the analyzer treats as having some overarching purpose or other—might reject this claim. But whatever the priest and deacon are disagreeing about, they are not disagreeing about the overarching purposes the parishioner's cognitive system. Thus, we can safely assume that the parishioner's functional states are held constant in relation to the two belief attributors.

[187] According to paramechanists, there is a single fact of the matter even if that fact is less than black and white: even if, for example, the fact is that the parishioner has a .55 credence that Jesus

both believes and disbelieves (Mandelbaum 2013) intersubjectively—in relation to every truth-tracking point of view. For paramechanists, idiosyncratic models of belief fail to capture what people actually believe precisely to the extent that they fail to describe intersubjectively determinate cognitive states of belief.

1.3. Rylean denials of intersubjective indeterminacy

Ryleans reject the type-identification of attitudes of belief with cognitive states individuated by their functional roles in cognitive systems. Nevertheless, Ryleans traditionally agree with paramechanists about the intersubjective determinacy of belief. Instead of achieving intersubjectivity by conflating attitudes of belief with cognitive states of belief, however, Ryleans have sought to achieve intersubjectivity by fixing attitudes of belief in relation to objective norms of interpretation.

In Section 2, I will explain how Davidson, Dennett, Baker, and Mölder all appeal to constitutive norms of interpretation in order to set intersubjective standards for veridical belief attribution. I will then argue against this Rylean tradition. On my view, Ryleans have failed to make a convincing case that there are constitutive normatively forceful norms of belief attribution, and should admit that what the parishioner believes might well be intersubjectively indeterminate: it might be that both the deacon and

---

rose again, or that the parishioner has a pro-resurrection belief fragment and an anti-resurrection belief fragment which play distinct functional roles. Indeed, even if it is objectively indeterminate what the parishioner believes—maybe cognitive states of belief are metaphysically vague—paramechanists would hold that it is still intersubjectively determinate: it is metaphysically vague from every truth-tracking point of view.

priest attribute belief veridically. Models of belief fix veridicality standards for belief attribution without relying on constitutive norms of interpretation. The mere fact that some models of belief better serve belief attributors' purposes than others does not suffice to establish that the normatively worse models fail to outer-recognize beliefs—that is, fail to be models *of belief*.

In Section 3, I will marshal empirical evidence that belief attribution practices vary across cultures and rely on a rich diversity of contingent norms of interpretation. In Section 4, I will use these findings to undermine traditional Rylean attempts to achieve intersubjectively stable, normatively forceful, constitutive standards of veridicality, and argue that beliefs are instead fixed relative to the idiosyncratic and normatively flawed models wielded by actual, individual belief attributors. In Section 5, I will argue that it is okay—that it threatens neither the representational nature of belief nor the empirical study of attitudes of belief—that practically significant intersubjective indeterminacy sometimes ensues. I will conclude by providing an example of how, in emphasizing the diversity that marks both believers and models of belief, my nonnormative and relativistic version of Ryleanism illuminates the scientific study of how folks, as understood by other folks, represent the world.

2. Interpretivisms and norms

Donald Davidson wrote that his anomalous monism led him (biographically speaking) to "the recognition of an irreducibly normative element in all attributions of

attitude" (2001a: 241). However, some scholars deny that the details of Davidson's interpretivism sustain his earnest claim to have offered a truly normative theory of belief. If these scholars were right, then my nonnormative Ryleanism would keep illustrious company after all. As it happens, these scholars have misinterpreted Davidson. Serendipitously, however, their misinterpretation can be co-opted by a relativistic Ryleanism. Timothy Schroeder's distinction between normative categorization schemes and normative force-makers will help illuminate the path towards a genuinely nonnormative interpretivism that relativizes beliefs to particular models of belief wielded by particular belief attributors.

2.1. Davidson on intelligibility as the norm of belief attribution

Schroeder has written an influential article titled "Donald Davidson's Theory of Mind is Non-Normative." Schroeder's central argument hinges on the assertion that although Davidson took believers to be (necessarily) rational, he did not take it to be constitutive of belief that believers *ought* to be rational. In Schroeder's (2003: 1) terminology, Davidson's theory of mind boasts a normative "categorization scheme"—beliefs are categorized as mostly true, rational, and coherent—but it lacks a normative "force-maker"—there is nothing intrinsic to Davidson's theory of mind that gives truth, rationality, or coherence their normative oomph. In other words, a presumption of a rationality is built into the Davidsonian conception of belief, but normatively forceful reasons to be rational are not. Nothing in Davidson's theory of mind makes rationality

inherently good.

For the sake of argument, I will grant that Schroeder has these details right; I will grant that Davidson's theory of mind does not include a force-maker that renders truth, rationality, or coherence inherently good.[188] Davidson did not explain how norms of *belief* acquire normative force.

However, Davidson did explain how norms of *interpretation* acquire normative force. As mentioned, Davidson took there to be "an irreducibly normative element in all attributions of attitude" (2001a: 241). Moreover, being a Rylean about attitudes of belief, Davidson took beliefs to exist only in relation to veridical interpretations. In particular, he argued that what somebody believes is fixed by the most charitable interpretation of their behavior: that is, the possible interpretation that would render the believer most intelligible as a rational agent (2001c: 215). This intelligibility norm boasts both a categorization scheme—belief attributions are ranked in terms of how intelligible they render believers—and a normative force-maker—the irreducibly normative structure of the communicative practice of triangulation. On Davidson's view, the very act of belief attribution presupposes that interpreters aim to render believers intersubjectively intelligible. This built-in aim imbues the intelligibility norm of interpretation with normative oomph. Together with Davidson's interpretivism, it ensures that beliefs are

[188] Verheggen (2016) provides reason to think, contra Schroeder, that Davidson's theory of mind does include a normative force-maker, though she admits that the constitutive normativity of belief is hypothetical rather than categorical. Glüer (2000) and Engel (2008) provide discussions that are friendlier to Schroeder's interpretation.

partly normatively constituted.

As discussed in Chapter 1, Davidson argued that problems of indeterminacy in belief attribution are solved via communicative triangulation, in which two radical interpreters openly reflect on their own beliefs, the other's beliefs, and the objects in the world that both sets of belief are about. Communicating with other belief attributors "forces us to the idea of an objective, public truth" (2001b: 170), "for the triangulation which is essential to thought requires that those in communication recognize that they occupy positions in a shared world" (2001c: 213). With this general notion of objectivity in hand, interpreters are then able to triangulate on particular objects in their shared world, including each other's beliefs. As Davidson wrote, "gauging the thoughts of others requires that I live in the same world with them, sharing many reactions to its major features" (220). Interpreters like the priest and deacon are able to argue about what the parishioner believes because, by being entrenched in the practice of triangulation from an early age, they both recognize that they occupy a shared world with intersubjectively stable properties. In particular, when the priest and deacon triangulate on the parishioner's beliefs, they are forced to the idea that there is an objective, determinate, empirically discernable fact about what she believes. The goal of belief attribution is to locate this fact.

Now, according to my telling of the parable, the priest and deacon seem to have located different—indeed, incompatible—facts. Davidson would chalk this disagreement up to either a lack of evidence or a failure of interpretation (on at least one

of their parts). However, in order to so much as qualify as Davidsonian belief attributors—even unskilled Davidsonian belief attributors—they must agree that there is an empirically discernable truth about what the parishioner believes, and that this truth aligns with the most charitable interpretation of the parishioner's behavior. They must agree that somebody is mistaken.[189]

The process of triangulation thereby gives normative force to the interpretation that would render the parishioner maximally intelligible to the priest and deacon. Indeed, according to Davidson, what triangulators believe cannot be articulated without reference to this ideal interpretation. Insofar as they are belief attributors at all, the priest and deacon both ought to interpret the parishioner in a particular, maximally charitable manner, and what the parishioner believes is determined by how her church leaders (and everybody else) ought to interpret her. Thus, contra Schroeder, Davidson's theory of mind is normative, not because what somebody believes depends on how they ought to believe, but because what somebody believes depends on how belief attributors ought to interpret them as believing.[190]

Schroeder anticipates this objection.

First objection: Davidson's theory of mind is genuinely normative, because there

---

[189] For example, they must both accept the conditional that if the priest understands his parishioner (and the nature of religious belief) as charitably as possible, then the parishioner objectively believes in the resurrection.

[190] Schroeder also overlooks Davidson's view that it is precisely this norm of intelligibility that makes it such that most beliefs must be true, rational, or consistent (2001c: 215).

> is a right way to interpret organisms and a wrong way to do so according to the theory. More subtly, some interpretations are better and some are worse; since the theory of mind is (in part) a theory of interpretation, normativity enters the theory here, at least. (Schroeder 2003: 10)

Schroeder responds as follows.

> Response: There is a right way and a wrong way to ascribe any property. There are better and worse interpretations of radio signals as being caused by quasars, for example; it does not follow that there is something especially normative about the domain of study of radio astronomy. Likewise, the fact that there are better and worse interpretations of the propositional attitudes of organisms in no way implies that propositional attitudes themselves are normative entities. All it implies is that theorizing (about anything) is a normative enterprise, which is a point in the philosophy of science, perhaps, but not necessarily one of significance to the philosophy of mind. (10)

However, Schroeder's response does not faithfully represent Davidson's view. Davidson argued against precisely this sort of equation of belief attribution and physical property attribution. In early work on the principle of charity, he explained that "each interpretation and attribution of attitude is a move within a holistic theory, a theory

necessarily governed by concern for consistency and general coherence with the truth, and it is this that sets these theories forever apart from those that describe mindless objects, or describe objects as mindless" (2001b: 154). The attribution of a belief differs from the attribution of a physical property insofar as it is *necessarily* governed by the principle of charity. Upon developing his notion of triangulation, Davidson doubled down on this stance.

> It is here that the irreducible difference between mental concepts and physical concepts begins to emerge: the former, at least insofar as they are intentional in nature, require the interpreter to consider how best to render the creature being interpreted intelligible, that is, as a creature endowed with reason. As a consequence, an interpreter must separate meaning from opinion [read: must separate veridical belief attributions from unveridical belief attributions] partly on normative grounds by deciding what, from his point of view, maximizes intelligibility. (2001c: 215)

No such requirement *necessarily* governs the attribution of physical properties. Physical property attributions *may* be normatively evaluated. Successful science depends on normative principles. Physical property attributions may even, contingently, always be guided by norms. Schroeder is right that it does not follow that there is something especially normative about radio signals and quasars. But Davidson held beliefs to differ

from radio signals and quasars partly insofar as belief attributions do not just happen to be guided by norms. Instead, belief attributions are necessarily governed—they are belief attributions only if they are governed—by an intelligibility norm of interpretation.

I suspect that Schroeder fails to recognize this distinction between these two senses in which attribution can be a normative enterprise because he pays too little attention to the crux of interpretivism: the thesis that to believe is to be aptly interpretable as believing. Although Davidson held that interpreters come to *know* physical properties via triangulation (just like mental properties), he insisted that physical properties exist—subject to the laws of physics—whether or not interpreters discover them. The tree falls in the forest even if no radical interpreters are around to hear it. Belief, on the other hand, emerges in the world alongside interpretation. Davidson steadfastly held that "there is no propositional thought without communication" (2001c: 213), and that events and attitudes "are mental only as described" (2001a: 215). A creature who is not (yet) a believer may be disposed to think, feel, and behave in ways consonant with taking the world to be a certain way, but this pattern of dispositions is not a belief unless triangulators reify it as such. Davidson's theory of mind is normative because he takes beliefs, unlike radio signals and quasars, to metaphysically emerge—rather than simply be discovered—via the irreducibly normative practice of triangulation.[191]

---

[191] Indeed, insofar as Schroeder's imputation to Davidson of an equivocation of physical states and mental states holds any water, it does so by highlighting Davidson's occasional flirtations with the idea that all truth is relative to interpretation. (A viable transposition of Davidson's

2.2. Dennett on prediction as the norm of belief attribution

Daniel Dennett also considers beliefs to emerge in relation to a constitutive norm of interpretation. For Dennett, what somebody believes depends on the most usefully and voluminously predictive interpretation of their behavior (1987: 15), whether or not that interpretation renders the believer maximally intelligible. Dennett (1987: 343) does take a baseline "Assumption of Rationality" (rather than a "Principle of Charity") to form a backdrop for all successful belief attribution, but only because evolution has designed the intentional stance to predict the behavior of *rational* creatures. In other words, Dennett agrees with Davidson that a presupposition of rationality is required for predictive success, but maintains that interpretations are (non)veridical relative to the predictive power derived from rendering believers intelligible (rather than relative to how intelligible per se they render believers).

Dennett admits that

> Prediction isn't the only thing we care about, of course. Folk psychology helps us understand and empathize with others, organize our memories, interpret our emotions, and flavor our vision in a thousand ways. (1998: 97-98)

However, Dennett stresses that belief attribution is capable of serving these various ends

---

(1974) claim that there is no such thing as a conceptual scheme is 'that there is just one conceptual scheme'.) But if this were the case, then Davidson would have to hold that physical states are constitutively normative, not deny that mental states are constitutively normative.

in virtue of its predictive power.

> Without its predictive power, we could have no interpersonal projects or relations at all; human activity would be just so much Brownian motion; we would be baffling ciphers to each other and to ourselves—we could not even conceptualize our own flailings. In what follows I will concentrate always on folk-psychological prediction, not because I make the mistake of ignoring all the other interests we have in people aside from making bets on what they will do next, but because I claim that our power to *interpret* the actions of others depends on our power—seldomly explicitly exercised—to predict them. (1998: 98)[192]

Dennett's predictability norm of interpretation boasts its own, Darwinian force-maker. The intentional stance has the teleofunction of enabling humans to predict—and only thereby interpret—behavior. People capable of adopting the intentional stance necessarily ought to provide the most predictive belief attributions because the intentional stance was designed (by evolution) to provide interpretations that are more usefully and voluminously predictive of intentional behaviors than interpretations from the design and physical stances. Thus, for Dennett, if the priest is able to predict more of his parishioners' behaviors than the deacon, over the entire course of their respective

---

[192] Davidson, for his part, criticized Dennett's emphasis on "the issue of prediction [a]s something of a red herring" (2001c: 81). I hash out this debate between Davidson and Dennett while defending my own alternative, below.

ministries, then the priest's intentional strategy is to be preferred to the deacon's, and his belief attribution is therefore more apt. If the priest's intentional strategy is maximally usefully and voluminously predictive, then the parishioner objectively believes that Jesus rose from the dead.

Dennett and Davidson posit different governing norms of interpretation in part because they disagree about whether it can be fully indeterminate what somebody believes.[193] Dennett allows that "there could be two interpretation schemes that were reliable and compact predictors over the long run, but which nevertheless disagreed on crucial cases" (1998: 117). In these (conceivable but practically irrelevant)[194] cases, belief is fully indeterminate. If the priest and deacon are equally good at predicting the parishioner's behavior over the long haul, then there is no determinate fact of the matter about whether she believes in the resurrection: "no deeper fact of the matter could establish that one was a description of the individual's *real* beliefs and the other not" (118). Or, to be more precise, there is one fact of the matter about what the parishioner believes in relation to the priest's intentional stance, and another, incommensurable fact of the matter in relation to the deacon's intentional stance. Both of these facts are perfectly objective, as well as (in principle) publicly accessible via empirical study of the priest and deacon's respective intentional stances, as well as the parishioner's thoughts,

---

[193] In Quine's language, Dennett allows for the possibility of "full indeterminacy", whereas Davidson only allows for the possibility of full indeterminacy's "trivial and indisputable cousin", "weak indeterminacy" or "ontological relativity" (Quine 1992: 50).

[194] Dennett stresses severally that "the bogey of radically different interpretations with equal warrant is … metaphysically important … but practically negligible" (1987: 29).

feelings, and behaviors. Nevertheless, Dennettian intersubjective indeterminacy arises only in extremely rare—if not apocryphal and merely conceivable—cases of competing, substantively different yet equally predictively ideal intentional strategies. For Dennett, the predictability norm of interpretation provides a clear veridicality standard for belief attribution, despite the fact that it allows for the possibility of intersubjective indeterminacy.

On Davidson's view, competing ideal interpretations that result in full indeterminacy are inconceivable. Following Quine (1992: 50), Davidson contrasts "full indeterminacy" with the sort of "weak indeterminacy" between whether the freezing point of water is 32°F or 0°C. Just as there is a single correct answer to the question whether or not the temperature is below freezing, no matter what temperature scale the measurer employs, "there is a correct answer to the question whether or not someone has a certain attitude" (2001c: 82), no matter what interpretive strategy the belief attributor employs. In triangulating, the priest and deacon must agree that there is a publicly accessible truth—fixed by the maximally charitable interpretation—about what the parishioner believes. They just disagree, temporarily, about what that truth is. Davidson wrote that "Dennett has urged that the answer to the [question whether there are objective grounds for choosing among conflicting belief attributions] is that there are no such grounds; but I do not think he has given any reason to accept this answer" (2001c: 82). Davidson rejected Dennett's hypothesis that two competing belief attributions could (even conceivably) be equally veridical. Even were it conceivable that

two competing intentional stances could be equally predictive, Davidson denied that it would be conceivable that they could render believers equally intelligible. On the contrary, the social practice of triangulation relies on the factive conviction, on the part of all triangulators, that if their disagreement is substantive, then one of them must be wrong. Every substantive difference between Davidsonian belief attributions corresponds to a difference in the intelligibility of the believer. For Davidson, then, the intelligibility norm of interpretation provides a monolithic veridicality standard for belief attribution, thereby dissolving the bogey of full indeterminacy.

2.3. Mölder on community standards as norms of belief attribution

Meanwhile, according to Bruno Mölder, along with the dispositionalists Lynne Baker (1995: Chapter 8)[195] and, in some moods, Eric Schwitzgebel (2013: 80), community norms set the veridicality standards for belief attribution. As these Ryleans plausibly argue, there is an objective (if abstract) fact of the matter about how most people in a community construe any given belief; most belief attributors wield roughly similar general-purpose stereotypes for most beliefs, such that sociologists could construct an abstract community-level general-purpose stereotype for any given belief that *p*. Mölder and Baker consider these community standards of belief attribution to constitute both

[195] Mölder (2010: 151–158) acknowledges a deep intellectual debt to Baker's notion of the "common-sense conception". Chapter 1 discussed Baker's version of the view that community norms fix veridicality standards of belief attribution. I focus primarily on Mölder in this chapter, since his view is laid out clearly and does not differ substantially from Baker's in this respect.

normatively innocuous *statistical* norms and normatively forceful *prescriptive* norms (Antony 2013: 532): they capture both how most members of a community *do* construe any given belief and, *therefore*, how the members of that community *ought* to construe that belief.

In particular, Mölder argues that a believer possesses a belief if and only if said belief is "canonically ascribable" to said believer. Veridical belief attributions are canonical belief attributions. Mölder (rightly, in my view) rejects Davidson and Dennett's "thought that perhaps we can specify the canonicality condition through the notion of an ideal ascriber" or ideal ascription, because "it makes all of our ordinary ascriptions noncanonical [and thus nonveridical], for ordinary folk cannot instantiate the ideal ascriber" (2010: 172). However, Mölder also preemptively rejects my alternative proposal that beliefs exist in relation to each arbitrary belief attributor's model thereof, on the grounds that a relativistic interpretivism would fail to provide a clear, intersubjectively discernable veridicality standard for belief attribution. Instead, Mölder insists that in order to set a veridicality standard that actual belief attributors can meet, "we need to locate the canonicality in the space between an arbitrary and an ideal ascription" (173).

Enter community norms. According to Mölder, "what is required for the application of mental terms [like belief] lies out in the open, in our common-sense psychology and it can be mastered by anyone who masters folk psychology" (2010: 2). On Mölder's view, attributions of belief are canonical—and thus veridical—insofar as

they would have been made by people with access to full evidence about believers and an ordinary mastery of folk psychology. In particular, Mölder identifies canonical ascriptions with attributions that ought not be revised in light of (1) evidence about the believers in question and (2) "facts about how ordinary ascribers interpret" believers (178). Actual belief attributors can ensure that an attribution meets the latter condition by ascertaining that competent members of their folk psychological community lack "warranted objections" to the attribution (174).

Mölder thus claims that community norms of belief attribution partly determine the possession of belief. To decide whether the priest or deacon interprets the parishioner correctly (given that they both possess as much evidence as practicable), we just have to figure out which man of faith interprets her in line with the community standard general-purpose stereotype for belief in the resurrection. Neither Mölder nor Baker (from whom he borrows the pertinent notion of community) specify how to determine the community in question—is it the congregation? the diocese? the Roman Catholic Church? Western culture? the set of all folk psychologists in the Universe?—though Mölder abstractly notes that canonicality is always indexed to a particular situation, and that "the features of that situation determine what sort of information is relevant and which revisions may be needed" (174).

On Mölder's view, "the range of ascriptions that are coherent and do not require revision can be quite large" in any given situation (175), but not so large as to result in intersubjective indeterminacy. Instead, he avers that "the fact that the canonical ascriber

is faced with two different or incompatible ascriptions is a good reason for seeking further revisions" (Mölder 2010: 184). Thus, he would argue that the fact that the priest and deacon are arguing is reason enough to assume that one of them falls short of the community norm for attributing belief in the resurrection. For Mölder, the parishioner (objectively and intersubjectively) believes whatever she would be interpreted as believing by a normal interpreter with access to all of the accessible evidence about her thoughts, feelings, and overt behaviors.

3. Is there a normatively forceful constitutive norm of belief attribution?

Mölder, Baker, Dennett, and Davidson all introduce normativity into their Rylean theories of mind in the guise of veridicality standards for belief attribution. Their shared insinuation is that the norms of interpretation they respectively introduce are theoretically crucial because they reveal how interpreters can be wrong about what people believe. As Mölder puts it, they reveal "the is/seems difference" that renders attributions of belief veridical or nonveridical (2010: 170).

3.1. Model-theoretic Ryleanism and the notion of an appropriate match

However, Ryleans need not countenance normative force-makers (in Schroeder's sense) as partly constitutive of belief in order to count some attributions as veridical and others as nonveridical. Consider the account of models of belief developed in Chapter 6. Somebody believes that *p* just in case they appropriately match a belief attributor's

general-purpose stereotype of that belief—that is, the generic pattern of dispositions to act, react, think, and feel that an interpreter takes to subsume all instances of believers taking *p* to be true.[196] This modified Schwitzgebelian account of models of beliefs provides Ryleans with a genuinely non-normative parameter for judging the veridicality of belief attributions, and thus the tools to construct a genuinely nonnormative Ryleanism.[197] On my view, a belief attribution is veridical just in case the believer really does possess a pattern of dispositions that appropriately matches the model of belief applied by the belief attributor.

One might suspect that the notion of an 'appropriate match' at play in this definition is normatively loaded. Davidson, Dennett, Baker, or Mölder might ask: does requiring appropriate matches to particular models of belief—like requiring intelligibility, or predictive power, or fulfillment of community standards—not amount to constitutive interpretive normativity?

It does not. As Schroeder would point out, the apparently normative term 'appropriate match' provides a normative categorization scheme for belief attribution but does not provide a normative force-maker. Some matches of patterns of dispositions to models of belief are categorized as appropriate, rendering belief attributions veridical;

---

[196] Likewise, somebody believes that *p* in a particular style just in case they appropriately match a belief attributor's style-specific model of that stylized belief.

[197] I avoid attributing the relevant view to Schwitzgebel himself because (a) it is unclear whether he takes the notion of an appropriate match to be normatively forceful and (b) Schwitzgebel sometimes goes in for constitutive community norms of interpretation. In the next few paragraphs, I argue that the Schwitzgebelian interpretivist *need* not embrace a constitutively normative theory of mind, rather than suggesting that Schwitzgebel actually does not.

others are categorized as inappropriate, rendering belief attributions nonveridical. However, there is nothing inherently good about attributions that appropriately match believers to models of belief. The fact that a believer lives in a pattern that appropriately matches a model of belief entails nothing beyond the fact that the believer has the modeled belief. In particular, it entails nothing about how the belief attribution fares in light of any norms with prescriptive oomph: a veridical interpretation does not *necessarily* render the believer the slightest bit more intelligible, more predictable, or more normal from an intersubjective point of view. Suitably altered, Schwitzgebel's conceptual machinery thus provides Ryleans with the tools for a genuinely nonnormative account of belief.

On the traditional Rylean views canvassed above, beliefs exist relative to these inherently normatively forceful norms of interpretation. For Davidson, beliefs exist relative to the intelligibility norm, which all belief attributors ought to meet insofar as they are belief attributors at all, because it is the heart of the communicative practice of triangulation from which beliefs emerge. For Dennett, beliefs exist relative to the predictability norm, which all belief attributors ought to meet insofar as they are belief attributors, because prediction is the proper function of the intentional stance. For Mölder, beliefs exist relative to community norms, which all belief attributors ought to meet insofar as they are belief attributors, because they undergird the intersubjective nature of folk psychology. Unlike Mölder, Baker, Dennett, and Davidson, I deny that belief attribution is constitutively normative, in any sense more interesting—more

normatively forceful—than the sense in which attributions can be sorted according to the categorization scheme of (in)appropriateness of match to a model of belief (and attendant (non)veridicality).

The priest and deacon wield substantively different models of the belief that Jesus rose again: they associate different dispositions with that belief, and perhaps differ on which degrees and respects of match to their models are sufficient to constitute a belief in the resurrection. As such, they interpret the parishioner differently, even though (*ex hypothesi*) they agree entirely about how she is disposed to think, feel, act, and react.

3.2. Diversity in models of belief

I suspect there is much more diversity in models of belief than other Ryleans are willing to allow. The cognitive capacity to attribute beliefs is (nearly) universal among humans, but there is robust evidence of significant differences in the developmental timeline of this cognitive capacity across cultures (Mayer et al. 2013; Slaughter & Perez-Zapata 2015). More to the point, Cecilia Heyes and Chris Frith (2014) have compellingly argued that ethnographic and experimental research on explicit mindreading tutelage (Pyers et al. 2009; Taumoepeau & Ruffman 2006 and 2008; O'Brien et al. 2011) reveals that culture has a strong effect on how humans learn to construe beliefs.

Heyes and Frith's case is further bolstered by evidence of cross-cultural variation in mental state vocabulary (Lillard 1998; Lomas 2016), the relationship between

attributions of belief and the cultural identities of believers (Perez-Zapata et al 2016), and divergent cultural predilections for perspective-taking (Wu & Keysar 2014). Members of collectivist cultures are less likely to attribute abnormal beliefs to their compatriots than members of individualist cultures, and members of cultures with few mental state terms appear to wield coarser models of beliefs than members of cultures with many mental state terms. There is also robust evidence that, from infancy, people are much more likely to imitate (and otherwise attend to the social cognitive teachings of) in-group members than out-group members (Kinzler et al. 2007; Buttelmann et al. 2013; Gruber et al. 2017). In short, it is plausible that the models—and judgments of appropriateness of match—that determine beliefs vary from cultural group to cultural group.

There is likely variance within cultures too. Shannon Spaulding (2016, forthcoming) has recently argued that "deep mindreading disagreements [are] common among neurotypical adults" (forthcoming: 10); referencing the Eric Garner case, Spaulding points out that "despite having access to the same evidence, intelligent, rational, well-meaning people profoundly disagree about how to interpret" believers (forthcoming: 5). Spaulding puts these disagreements down to a panoply of factors, several of which have to do with how different interpreters model beliefs (and other attitudes). Our believer-specific models of the beliefs of in-group members are more

nuanced than our models of the beliefs of out-group members (Haslam 2006).[198] We also select different general-purpose stereotypes when interpreting in-group versus out-group members as believers (Krueger 1998; Ames 2004a, 2004b). In quotidian, practical contexts, lay belief attributors from different subcultures are prone to model the beliefs of a particular believer differently. As Spaulding writes, "this is not simply a matter of theoretical discussion not capturing the messy, empirical details. Individuals with different social backgrounds generate different mindreading judgments in predictable ways" (forthcoming: 26).[199]

This evidence coheres well with a developing consensus that a rich diversity of norms govern lay belief attribution. Davidson presupposes that people attribute beliefs in order to render believers intelligible, Dennett presupposes that people attribute beliefs in order to predict behaviors, and Baker and Mölder presuppose that people attribute beliefs in order to explain behaviors in ways that are decipherable by the folk

---

[198] This evidence points to intracultural intersubjective indeterminacy because in-group membership is highly context specific, turning on things like shared hobbies as well as cultural identities (Tajfel 1974).

[199] Even individuals with similar backgrounds may operate with substantively different models of belief. It might be, for example, that philosophical theorizing has had a top-down influence on how philosophers like Schwitzgebel (2002) and Tamar Gendler (2008) construe beliefs. For example, Schwitzgebel may construe Juliet (the implicitly racist teacher) as believing—in-between believing, yes, but still fitting the general-purpose well enough to count as a believer—that black students are intellectually inferior to white students. Gendler, on the other hand, construes Juliet as *believing* that black students and white students are intellectual equals, but *alieving* that they are not. Of course, it is perfectly possible that philosophers' theorizing (along with other cultural activities) does not have such vivid top-down impact on their folk craft of mindreading. (If social psychologists were to give up the assumption that attitudes of belief are type-identical to cognitive states of belief, then they might be more inclined to study whether cultural activities like philosophizing strongly influence how belief attributors construe beliefs.)

psychological community. They are each right, sometimes. Belief attributions function not only to render believers intelligible and predict and explain behavior, but also to contextualize (Tanney 2013), judge (Morton 2003; Monroe & Malle 2017), regulate (McGeer 2007; Hrdy 2009; Andrews 2015), and manage impressions of (Malle, Knobe & Nelson 2007) how believers lead their lives. It is far from obvious that these diverse functions are always best fulfilled by first rendering believers maximally intelligible, or maximally predictable, or maximally explanatory to normal folks.

All Ryleans acknowledge that what we count as a good—and not merely veridical—belief attribution depends on what norms are at play in the context of attribution. I want to suggest that there is no reason to expect that belief attributors always identify the exact same patterns of dispositions with beliefs when they aim to predict behavior as when they aim to look cool in front of their friends. Moreover, there is no reason to expect that belief attributors always *ought to* identify the exact same patterns of dispositions with beliefs in different normative contexts. As Spaulding writes, when "individuals have different goals—as is often the case in real world social interpretation—not only will they take different information as input, they will process that information very differently as well" (forthcoming: 33). In some normative contexts—like when we are trying to express a nuanced, personal understanding of a loved one—slower processing that exploits believer-specific models (and perhaps even constructs them on the fly) is preferable; in other normative contexts—like when we are trying to put a suspicious stranger in their place—faster processing that exploits general-

purpose stereotypes is preferable.

Perhaps there are epistemological or ethical norms that happen to seriously constrain how belief attributors ought to model certain beliefs across contexts, given the belief attributors' general epistemological and ethical ends. But unlike traditional interpretivists, I see no reason to take these norms to be constitutive of belief attribution. (Perhaps they are constitutive of rational agency, or decent personhood.) Thus, when I deny that belief attributors always ought to wield particular models across normative contexts, I mean to deny only that such a norm is built into the metaphysics of belief attribution (and thus built into the interpretivist metaphysics of belief).[200]

If models of belief that fix the veridicality standards for belief attribution can vary across normative contexts, then Davidson, Dennett, Baker, and Mölder are wrong to hold that what somebody believes is always determined by any single, privileged norm of interpretation. Neither paramechanical views nor traditional Ryleanisms have the flexibility to countenance widespread, practically significant variation between how competing belief attributors veridically construe beliefs. If such variation is rampant—because of (inter- and intra-cultural) differences in general-purpose models of belief or differences in the goal-specific models of belief deployed by attributors working towards different normative ends—then Ryleans should study and embrace this variation rather than attempting to explain it away.

---

[200] I will return to the defense of this point in sections 4 and 5.

3.3. Variation in belief between belief attributors

My relativistic version of Ryleanism countenances the possibility of widespread and practically significant variation, between belief attributors, in attitudes of belief. On my view, each and every belief exists in relation to a particular model wielded by a particular belief attributor.

Consider, once again, the analogy with relationalist views of color. If two different color perceivers—myself and a black bear, say—stably perceive an object as colored in two different ways, then there is no intersubjective standard to appeal to in order to prove that one of us has misperceived the surface of the object. On the contrary (especially if we have both seen the object in many lights and from any vantage points), we have respectively perceived two different relational properties of the object: the color the object truly has in relation to my perceptual system, and the color the object truly has in relation to the bear's perceptual system. These two perceiver-relative colors are nevertheless both objective and (at least setting aside the problem of other minds) intersubjectively discernable. I can come to appreciate that the bear veridically perceives the color of the object, even though she perceives it as differently colored than it is in relation to my own perceptual capacities.

Analogously, there is no intersubjective standard that could probatively settle the debate between the deacon and the priest about what the parishioner believes. By hypothesis, they are both deeply familiar with the parishioner's dispositions. They differ only on which patterns of dispositions they identify with the belief that Jesus rose again.

The two men make sense of the parishioner's thought and feelings, her actions and reactions, in different ways. She has the belief that Jesus rose again for the priest, because that is how the priest understands mental profiles like her's. She lacks that belief for the deacon for the same reason. There is no constitutive norm of interpretation—no way in which the two men both ought to understand the parishioner's mental profile *in order to count as belief attributors*—to appeal to in order to decide between the two interpretations.

The parishioner's two attributor-relative beliefs are nevertheless both objective and intersubjectively discernable. Without revising their models of belief, the priest and deacon will (rightly) refrain from agreeing about what the parishioner believes: she really does believe one thing for the priest and another for the deacon. Nevertheless, the two men may well come to agree about the fact *that* the parishioner believes one thing for the priest and another for the deacon. Through long conversation, the priest may come to appreciate—if not adopt—the deacon's old-fashioned, by-the-books stance on faith, and thereby come to appreciate that the parishioner lacks belief in the resurrection in relation to the deacon's model. By achieving this understanding of the deacon as a belief attributor, the priest can even come to see the deacon's belief attribution as veridical. It *is* veridical, given that the deacon has an accurate grasp of the parishioner's dispositions and has ascertained that she fails to appropriately match his model of belief in the resurrection. Even in this harmonious scenario, it is fully intersubjectively indeterminate what the parishioner believes, given that the priest stands by his new-

fangled practice-first model of belief in the resurrection.[201]

This claim—that there can be substantial variations in belief between belief attributors—is less radical than it may first appear, for a couple of reasons.

First, practically significant variation arises only in idiosyncratic cases. Just as everybody sees the viral dress as (subtly different shades of) blue and black in most lights,[202] in most scenarios belief attributors differ only subtly on which patterns of dispositions they associate with any given belief, and thus avoid talking past one another. As discussed in Chapters 4 and 5, strong evolutionary and societal pressures ensure that most humans model beliefs in normalized manners. Indeed, the priest and deacon will not only agree about the vast majority of what this particular parishioner believes; they will also agree about whether or not most other parishioners in the parish believe in the resurrection. Like the subtle differences in our perceptual systems, the subtle differences between the priest and deacon's respective models of belief only result in conflicting belief attributions in idiosyncratic cases.

Second, there are variations in style of belief between believers. Whether or not the parishioner believes in the resurrection varies from general-purpose model to general-purpose model. However, as argued in Chapter 6, belief attributors like the priest and deacon often go beyond general-purpose models and construct believer-

---

[201] This full intersubjective indeterminacy differs from the indeterminacy between whether the temperature is 32°F or 0°C because the priest and deacon's models of belief in the resurrection track substantially different patterns of dispositions, rather than merely representing the exact same pattern of dispositions on different scales.

[202] See Figures 6 and 7 in Chapter 4.2.1.

specific models that capture the particular styles in which believers take the world to be certain ways. Thus, variation in belief between belief attributors is sometimes only variation in the stereotypes through which belief attributors get an initial grasp on complex, stylized beliefs.

Of course, stereotypes matter. People frequently attribute beliefs using general-purpose models in order to classify—and thereby make practically significant judgments about—folks as believers that *p*. Nevertheless, clashing interpretations of what somebody believes do not necessarily involve differing reads on the details of how that person is disposed to act, react, think, or feel. They often just involve different models that recognize different patterns arising from the believer's full set of dispositions. If both the priest and deacon were to construct new believer-specific models of the parishioner's idiosyncratic attitude towards the resurrection, they might well construct functionally equivalent models, and agree that the parishioner has a skeptical-but-reverent attitude towards the resurrection. Even so, they would continue to wield incompatible general-purpose models of belief in the resurrection, and continue to assess what the parishioner believes differently (but equally veridically) in light of their respective general-purpose models.

4. A dilemma for Ryleans

If this is right, then Ryleans are left with two options. First, we can deny that norms of interpretation are constitutive of beliefs. Second, we can insist that norms of

interpretation play constitutive (as well as normative) roles in belief attribution, but allow that different norms reign over belief attributions made for different purposes. Either way, the possibility of intersubjective indeterminacy inevitably follows on Ryleanism's heels.

4.1. Horn one: rejecting constitutive norms of belief attribution

The first option, which I favor, is to deny that interpretivism must be a normative theory of belief. Davidson, Dennett, Baker, and Mölder all argue that norms must be constitutive of belief in order to provide veridicality standards for belief attribution. But models of belief, with their (non-normatively-forceful) veridicality standards, rid Ryleans of the need to mix the normative with the metaphysical in this fashion. Appeal to a normatively ideal or canonical attribution might dictate which models belief attributors ought to adopt, but it does not dictate which models belief attributors actually wield to veridically attribute beliefs—it does not dictate which models actually metaphysically determine beliefs.

Recall Schroeder's misinterpretation of Davidson.

> There is a right way and a wrong way to ascribe any property. There are better and worse interpretations of radio signals as being caused by quasars, for example; it does not follow that there is something especially normative about the domain of study of radio astronomy. Likewise, the fact that there are better

> and worse interpretations of the propositional attitudes of organisms in no way implies that propositional attitudes themselves are normative entities. (Schroeder 2003: 10)

Schroeder gets Davidson wrong, but for my money Schroeder's pseudoDavidson is right about the normativity of belief attribution. There are better and worse ways to model beliefs, given belief attributors' goals; it does not follow that there is something especially normative about belief attribution. On the contrary, my relativistic Ryleanism takes belief attribution to involve a normative categorization scheme with no constitutive normative force-maker.

Consider community standards (a la Baker and Mölder). In any given interpretive context, there may be objective facts of the matter about the patterns of dispositions that *most* people (in the parish, or diocese, or city, or civilization, or world) consider to be constitutive of any given belief. Or there may be facts about the general shapes of the models wielded by *normal* people, or *experts*. For example, perhaps the parishioner disbelieves in the resurrection, according the models of most—or the most normal, or most expert—members of the parish. Such community and expert standards, presumably developed partly on the basis of interpretive success, play crucially important roles in setting norms that causally influence the models of belief that members of the community (ought to) develop and adopt. But the fact that community (and expert) folk psychological standards play these crucially important *normative* and

causal roles in interpretive practices does not entail that they directly fix the *metaphysical* constitution of beliefs.

Of course, it is perfectly coherent to speak abstractly of the models of beliefs wielded by groups like Americans, or Catholics, or whomever. We can (and do) fruitfully talk about the common features of the general-purpose models wielded by most (or the most salient) members of a community. However, these group-level models are merely abstractions from (and heuristic stand-ins for) the idiosyncratic models wielded by the particular belief attributors collected by the group under discussion.[203] These abstractions are metaphysically derivative of the particular models from which they are abstracted. Individual belief attributors do often have strong pragmatic reasons to alter their models of belief to more closely approximate the models of belief that are normal in their communities. Moreover, folks often exert themselves to cause others to adopt models that approximate their own models, by way of social activities like teaching, praising, coercing, and teasing. However, this normative and causal pull towards homogeneity in belief attribution does not render recalcitrant outliers' abnormal belief attributions any less veridical (though it might render them significantly less practical when the outliers attempt to communicate with normie belief attributors).

Much the same response applies to Dennett and Davidson's respective appeals to the norms of intelligibility and predictability. There may be an objective fact of the

[203] If group-level models are emergent from particular attributors' models, they are transformationally (as opposed to synchronically) emergent (Humphreys 2016).

matter about which attribution of belief is most conducive to rendering believers maximally intelligible or predictable. Again, appeal to an ideal (strategy of) belief attribution might play an important role in dictating which models of belief people ought to adopt. But it does not dictate which models actually constitute beliefs in our nonideal world.[204]

The analogy with color relationalism remains informative. Maybe it would be ideal, practically speaking, for me to perceive the ripe wild raspberry as a more vibrant, brighter shade of red—even easier to spot against the green leaves. But I do not; the raspberry actually sports a dull hue in relation to my perceptual capacities. If the black bear perceives the raspberry as more vibrantly colored than I do, her color-perceptual point of view may be normatively superior to mine. She is better at spotting raspberries amongst leaves, because of her way of perceiving color. But this normative superiority does not make a metaphysical difference. The raspberry remains a dull shade of red relative to me, even if I would be better off were it more brightly colored (as it is relative to the bear). Neither the bear nor I descriptively misrepresent the color of the raspberry.

---

[204] If Ryleans must pick an overarching norm of interpretation, Davidson's choice is perspicuous. Some intelligibility stems from any useful interpretation, no matter what other norms are in play. But this does not vindicate Davidson's intelligibility norm of interpretation. First, Ryleans need not pick an overarching norm of interpretation; not all interpretations have to be useful in order to be veridical. Second, the notion that all belief attributions must provide *some* intelligibility does not entail that only the belief attribution that renders believers *maximally* intelligible is veridical. Third, Davidson makes an error in assuming that the intelligibility of believers as believers is something that can be fixed independently of the psychologies of particular belief attributors and the norms at play in the particular social contexts in which belief attribution occurs. One main thrust of my argument in this chapter is that just *how* people are intelligible as believers depends, in large part, on both belief attributors' models and contextually variable norms of interpretation.

Likewise, neither the priest nor deacon descriptively misrepresent what the parishioner believes, even if one of their models of belief in the resurrection allows for greater predictability or intelligibility.[205]

Now, it may be that the deacon ought (again, normatively speaking) to revise his general-purpose model of belief in the resurrection to render his belief attributions more useful (for fostering a sense of community in the parish, for instance). If carried out, this revision would make it such that the parishioner believes that Jesus rose again for the deacon. Nevertheless, it would not retroactively make it such that the parishioner believes according to the deacon's previous model. In other words, it would not make

---

[205] Reflection on an analogous debate in the metaphysics of race literature may be helpful. Several philosophers have converged on the view that disagreements about the reality of race hinge on a shared understanding of the empirical facts but differing metaphysical definitions of race (Taylor 2000; Mallon 2006; Haslanger 2012). The same point has been made in the literature on disability (Riddle 2013). If one defines race in terms of biological essences, then Kwame Anthony Appiah (1985) is right that race isn't real. If one defines race in terms of biological populations, then Quayshawn Spencer (2014) is right that race is real but that the metaphysics of race is largely irrelevant to questions of social justice. If one defines race in terms of a history of discrimination, then Du Bois (1897) is right that race is real and that the metaphysics of race is extremely relevant to questions of social justice. Paul Taylor, Ron Mallon, and Sally Haslanger have suggested that we should move beyond metaphysical discussions of how race is defined (and whether or not it is real on that definition) and argue instead about how we pragmatically *ought to* define race, given anti-racist ethical ends (as opposed to the purely epistemic principles supposedly guiding descriptive metaphysical inquiry). In other words, these philosophers all want to reorient the philosophical debate about how to construe race to center on the normative rather than metaphysical question. There is a lesson here for intraRylean disagreements about the metaphysics of belief. Unlike recent metaphysicians of race, Ryleans have failed to cleanly distinguish metaphysical questions and pragmatic questions. The parishioner believes for the priest, and disbelieves for the deacon (similarly: race is real for Du Bois, but not real for Appiah, since they model race differently). Another important and interesting—but distinct—question is pragmatic. *Ought* the deacon revise his general-purpose stereotype of belief in the resurrection, just as (I think) Appiah *ought* to revise his understanding of what would make race real? One can give a principled case for an answer to that normative question without assuming that, metaphysically speaking, belief attribution is irreducibly normative.

the deacon's previous belief attribution nonveridical. The deacon—having revised his model to match the priest's—is correct that the parishioner believes, and was previously—before revising his model—correct that the parishioner did not believe. This diachronic inconsistency is possible not because the parishioner changed (we can assume that all of her dispositions remained constant) but because the deacon changed. The parishioner really did not believe in light of the deacon's previous point of view, and really does believe in light of the deacon's revised point of view. The deacon's revised point of view may be normatively superior (given his practical ends), but it is no more veridical: the previous point of view made no descriptive errors.

Alternately, the deacon might come to appreciate the fact that the priest's model of belief is preferable than his own for some purposes, yet still refuse to alter his conservative model. (Perhaps it is a better model for the purpose of solidifying his identity as a God-fearing man, even while being a worse model for the purpose of shepherding the parishioner.) In that case, the deacon would still be right to describe the parishioner as disbelieving *for him*, even though he knows that it would have been better (for his immediate purposes in counseling the doubtful) if he had modeled belief differently.

4.2. Horn two: embracing a pluralistic normative Ryleanism

I will not press the point that normative standards play no constitutive role in the metaphysics of belief any further. My nonnormativism is not, strictly speaking, crucial to

my overall case for a relativistic Ryleanism about belief. For the Rylean's second option is to double down on the normativity of belief attribution, and argue that different norms constitutively and forcefully govern belief attributions made for different purposes.

How belief attributions are properly normatively evaluated depends on what goal(s) the attributions subserve. Thus, even if norms of interpretation are constitutive of belief, different belief attribution practices (governed by different norms) will constitute beliefs in different ways. On this view, the predictability norm sets veridicality standards when the goal of attribution is to predict behavior; but the attribution that renders believers most predictable is not veridical in every context. For example, it is not necessarily the veridical attribution when the belief attributor's goal is to regulate behavior. The deacon may wield the best model of belief in the resurrection for the purpose of prediction, whereas the priest wields the best model for the purpose of regulation. If the deacon and priest are at cross-purposes—perhaps the deacon wants to predict which pointed questions the parishioner will ask, whereas the priest wants to unite his parish in faith—then both of their attributions can be veridical, even stipulating that belief is determined with reference to normative ideals.

If normative standards play a constitutive as well as normative role, then this may be evidence of further, not less, variation in attitudes of belief. A normative relativistic Ryleanism would dictate that individual believers have different beliefs depending on what the attributions of belief are for—the most predictive belief

attribution will often come apart from the most regulative belief attribution—as well as depending on who is doing the attributing. I myself find this pluralism about constitutive norms implausible, if only because I see no reason to countenance constitutive norms at all. Many models of belief are general-purpose tools that people employ across a wide variety of normative practices. These tools are causally shaped by myriad norms—including community standards, intelligibility, and predictability—but, by my lights, metaphysically constituted by none. Nevertheless, if Ryleans persist in taking normatively forceful norms of interpretation to be constitutive of beliefs, then they must grapple with the normative complexity inherent in actual belief attribution practices.

4.3. Relativistic Ryleanism and intersubjective indeterminacy

Insofar as the normative and the metaphysical collide, belief is determined relative to the diverse goals of our belief-ascribing practices. Insofar as the normative and the metaphysical should be kept distinct, belief is determined relative to the diverse conceptions of belief inherent in our belief attribution practices. Either way, beliefs are constituted in relation to each of the standards by which actual people attribute beliefs. Relativistic Ryleanism is the only extant theory of belief with the built-in flexibility to countenance the multitudinous messy ways in which idiosyncratic belief attributors, across cultures and normative contexts, get a grip on idiosyncratic believers. Like Dennett but unlike Davidson, Baker, and Mölder, the relativistic Rylean must embrace

the full intersubjective indeterminacy of belief. Unlike even Dennett, the relativistic Rylean must also embrace the possibility that full intersubjective indeterminacy is practically significant. For example, the indeterminacy about what the parishioner believes make a substantive difference to the counsel provided by the priest and deacon.[206]

5. The messy metaphysics of belief in practice

I have argued that normatively forceful constitutive norms of interpretation are not required to set veridicality standards for belief attribution. However, I have not yet addressed the other major reasons that traditional Ryleans have adopted their preferred norms. Among the desiderata for a theory of attitudes of belief are (1) that it sets beliefs apart from other attitudes, and other aspects of believers' personalities, by making sense of beliefs' distinctive properties (such as their representational nature and mind-to-world direction of fit), and (2) that it is usable by the natural and social sciences that investigate attitudes of belief. It might be thought that, by dispensing with constitutive norms, and by fixing beliefs relative to the particular models wielded by particular belief attributors, my relativistic Ryleanism falls short of achieving these desiderata. In particular, an objector might ask the following two pointed questions. First, if models of belief are causally shaped by a wide variety of norms, yet constituted by none, then in

[206] Spaulding (forthcoming) illustrates the potentially grave practical consequences of divergent models of belief by illuminating divergent interpretations of the murder of Eric Garner by the NYPD.

virtue of what do they remain models of *belief*? And second, if beliefs vary in practically significant manners between belief attributors, then how can social scientists hope to study beliefs at the population level? I will take these questions in turn.

5.1. What makes attributor-relative beliefs *beliefs*?

In developing their respective interpretivisms, Davidson, Dennett, and Mölder all emphasize that humans' very conception of others as minded depends on the attribution of belief. In Davidson's (1973) language, radical interpretation can only get off the ground if radical interpreters can reliably attribute "attitudes of holding true" to interpretees. According to Davidson, the principle of charity provides the only reliable means of tracking attitudes of holding true; Dennett thinks only the combination of an assumption of rationality and a focus on maximizing predictability will do; Mölder thinks aligning our belief attributions with those of others in our community does the trick. In any case, all three interpretivists agree that these norms serve not only to provide veridicality standards for belief attribution, but also to ensure that interpreters (who are necessarily guided by those norms) actually latch onto *beliefs*—attitudes of holding true—rather than other, less folk psychologically fundamental patterns of dispositions. Interpretees are rendered maximally intelligible (or predictable, or explainable to normal folks) only insofar as interpreters can get a grasp on how they represent the world.

By giving up on constitutive norms of interpretation, and allowing a wide range

of norms to causally shape models of belief, it might be objected that I have given up on what makes belief *belief*. Could a model largely shaped by norms that have very little to do with how the believer represents the world—aesthetic norms, for example, or norms of etiquette—really be a model of belief? If the priest's model were to dictate that somebody believes in the resurrection just in case they are disposed to root for the same sports team as the priest, refrain from laughing when the priest's toupee is askew, and so on, then in virtue of what would that count as a model of belief (as opposed to a model of comradeship or something)?

As it happens, people do not pervert their models of belief to such egregious degrees. Beliefs do not emerge relative to any old cognitive models; they emerge relative to the models of taking the world to be some way—models of Davidson's "attitudes of holding true"—which social psychologists have demonstrated to be almost universally wielded by modern humans. As discussed in Chapters 4 and 5, the human capacity to construct models of belief has co-evolved (and occurrently co-develops) with mindshaping practices—including imitation, pedagogy, conformity to norms, and narrative self-constitution—which serve to alter believers' dispositions to be more cleanly modellable (McGeer 2007; Sterelny 2012; Zawidzki 2013; Heyes 2018). Both mindshaping and mindreading practices are robustly (albeit contingently) normatively guided; students and teachers of folk psychology frequently aim to maximize their own and others' intelligibility, predictability, and normalcy (among other ends). When their models of belief fail to serve these ends, folks revise their models, often under strong

social pressure from their peers. Models of belief that have been overly causally influenced by aesthetic norms fail to serve belief attributors' ends. Indeed, models of belief serve most folk psychological practices best when they faithfully subsume dispositions to act, react, think, and feel in ways that fulfill believers' goals only if the world is actually the way that believers are being modeled as taking it to be. Ryleans have no need to posit constitutive norms of belief attribution because the norms that contingently but reliably guide belief attribution practices have phylogenetically and ontogenetically ensured that belief attributors are responsive to how believers represent the world.

All realist theories—Rylean or paramechanical—take attitudes of belief to be special in that they realize distinctive perspectives on the world, rather than literal representations—reproductions—of the actual properties of the world. My relativistic version of Ryleanism recognizes an additional special aspect of attitudes of belief: they realize distinctive perspectives on the world, *as appreciable from another distinctive perspective*. In other words, beliefs are believers' ways of representing the world, as modeled by particular belief attributors. By my reckoning, this theoretical convolution is made up for by relativistic Ryleanism's unique ability to pin down the slippery objects of everyday belief attribution practices.

For example, it is illuminating to grasp, in one synoptic picture, how the deacon and priest respectively model the parishioner's beliefs regarding the resurrection. By understanding both the deacon's perspective and the priest's perspective, we can come

to understand the parishioner's own distinctive perspective better. She lives both as if Jesus rose again (in relation to the priest's practice-oriented model of belief) and as if Jesus did not rise again (in relation to the deacon's creed-oriented model of belief). This conjunction of attributor-relative beliefs sets her perspective apart from those of most other people, who either believe or disbelieve in relation to both models. Without appreciating the priest and deacon as wielding fully-intersubjectively-indeterminate-yet-equally-viable models of belief, we would not be positioned to fully appreciate the details of the parishioner's distinctive doxastic relationship to the resurrection. Indeed, the priest might construct a singularly subtle and profoundly informative believer-specific model of the parishioner's unique style of belief (and the deacon might construct a similarly commendable believer-specific model of the parishioner's unique style of disbelief), precisely by taking into account the pattern of dispositions identified by the other church leader.

Here is another case. Most farmers in Mississippi, North Carolina, Texas, and Wisconsin deny that the climate is changing in a manner that will adversely affect crop yields. Nevertheless, these very same farmers are disposed to "diversify crops, buy crop insurance, modify lease arrangements, and exit farming" in direct response to the changing climate (Rejesus et al. 2013: 701). Do these farmers believe the climate is changing? Not from their own perspective (nor the perspectives of many social scientists, who measure belief according to assent to questions on opinion polls). These farmers have no use for understanding their own agricultural dispositions in terms of an

overarching belief about the increase of global mean temperatures. Nevertheless, these same farmers surely do believe in climate change from the perspective of a belief attributor who considers beliefs to be manifested primarily in practical behavior, rather than theoretical—much less political—speech.

5.2. Can relativized attitudes of belief be objects of scientific inquiry?

Dan Kahan has termed the general problem of scientifically studying what people like the United States farmers believe the "science-of-science-communication measurement problem" (2015: 1). If people reliably think and talk one way and act another—and thus are interpreted as disbelievers from one viable perspective and as believers from another viable perspective—then what are social scientists to say that they believe?

This question brings us to the general worry that if beliefs are relativized to particular models of belief, then they are scientifically intractable. As Kahan's measurement problem reveals, it is true that beliefs are difficult to study scientifically. Nevertheless, relativistic Ryleanism helps illuminate how scientists can make progress on the measurement problem (as well as other difficulties with scientifically studying belief). According to Ryleanism, understanding beliefs requires understanding belief attributors. In order to really understand the beliefs of members of any given society, social scientists must study how people in that society model beliefs (in addition to studying what people in the society believe in relation to the investigators' own

detached, anthropological models of belief).

Working under a relativistic Rylean framework, the task of social scientists of belief would be to reveal what people believe in relation to various salient models of belief. It would be an important sociological discovery—important, for example, for figuring out how to devise and pitch climate-friendly agricultural policy—to find that a significant proportion of US farmers believe in climate change in relation to the perspectives of members of the USDA, but disbelieve in climate change in relation to their own perspectives, as well as the perspectives of most House Republicans. To denounce relativism, and claim that the farmers *really* believe or disbelieve (or even in-between believe) whatever a privileged constitutive norm of interpretation dictates that they believe, would be to preclude a valuable source of insight into the ways in which attitudes of belief—as well as the practices of belief attribution from which they arise—are socially conditioned.

To conclude, Chapter 8 will discuss how my relativistic Ryleanism about belief can be put to good use in another science of the attitude of belief: the comparative psychology of mindreading.

6. Conclusion

Paramechanical theories of belief identify attitudes of belief with functional states (or their realizers) within the cognitive systems of believers. In this chapter, I have argued that a single believer—with a single cognitive system comprising a single set of

subpersonal computational states giving rise to a single set of personal dispositions fulfilling a single set of teleofunctions—can have significantly different attitudes of belief in relation to different belief attributors. If this is true, then paramechanical theories of belief—and, indeed, nonrelativistic Ryleanisms about belief—must be false. Only a relativistic Ryleanism captures the actual, messy diversity of veridically applicable models of belief.

Chapter 8: Brute believers and monkey mindreaders

1. Introduction

Humans are not the only believers in the world. Other animals have beliefs too. As of the time of this writing—8:28pm—my increasingly desperately meowing cats believe it is dinner time. (Unfortunately for everyone, their beliefs are false: cat dinner is not served until ten o'clock.) Analytic philosophers of mind have often struggled to account for the nature of animal belief, and some have denied the phenomenon outright. Most notoriously, Stephen Stich (1979) and Donald Davidson (1982) have argued that all believers are language users, and that to attribute beliefs to cats is crass anthropomorphism. This latter invective was also given voice by the majority of ethologists working in the 20th century, who opposed the attribution of any human-like psychological traits to nonhuman animals.

In this final chapter, I will apply the lessons of the preceding chapters to three ongoing debates about the philosophy and sciences of animal belief. First, ought anthropomorphism be prohibited in animal cognition research? Second, do nonhuman animals have beliefs? Third, are nonhuman animals belief attributors? Applying my relativistic Rylean account of belief to these methodological and substantive questions regarding animal cognition research will in turn illuminate a key moral of Chapters 1–7: mental differences matter.

2. Anthropomorphism

In attributing beliefs (not to mention desperation) to my cats, am I committing the pathetic fallacy? Relatedly, ought anthropomorphism—the attribution of human-like psychological characteristics to nonhuman animals—be prohibited in animal cognition research? My short answer is 'no, but it should be used wisely'. My long answer begins with a look back at the history of the controversy.

2.1. Morgan's Canon

The debate about the role of anthropomorphism in animal research has often revolved around the methodological dictum known as 'Morgan's Canon'. At the International Congress of Experimental Philosophy in 1892, the ethologist and philosopher C. Lloyd Morgan declared that "in no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be fairly interpreted as the outcome of the exercise of one which stands lower in the psychological scale" (Dixon 1892: 392; repeated in Morgan 1903: 242). At a first gloss, the Canon requires researchers to give the least mentalistic fair interpretation.

Putting his Canon into action, Morgan criticized George Romanes for inferring that dogs "have general ideas of 'good-for-eating' and 'not-good-for-eating,' *quite apart from any particular objects of which either of these qualities happens to be characteristic*" (Romanes 1888: 36). Morgan argued that all relevant doggish behavior can be explained by dogs having the capacity to recognize *particular* objects as 'good-for-eating.' This

recognition is, itself, the exercise of a psychical faculty; indeed, Morgan wrote that "this is a concept in [some uses] of the term, I admit" (Morgan 1891: 349). However, it the exercise of a psychical faculty that stands lower on the psychological scale—that is, it is a less sophisticated cognitive capacity—than the capacity to consider the concept of 'good-for-eating' in the abstract, apart from any particular edible objects. Since Morgan knew of no independent evidence bearing on the question "whether this quality can be isolated by the dog, and can exist in his mind divorced from the eatables which suggest it" (348), he was left "unable to attribute to the brute" the higher "power of analysis—the power of isolating qualities of objects" (350). It is possible that dogs have this higher power, but the less anthropomorphic interpretation was fair in light of the available evidence, so Morgan's Canon dictated that it won the day.

By the 1930s, Morgan's Canon was regularly understood as recommending a strict ban on the attribution of psychological capacities to animals; for example, B.F. Skinner wrote that while Darwin attributed mental states to animals, "Lloyd Morgan, with his law of parsimony, dispensed with them in a reasonably successful attempt to account for characteristic animal behavior without them" (1938: 4). By the end of the 20th century, Morgan's Canon had become "possibly the most important single sentence in the history of the study of animal behavior" (Galef 1996: 9). The importance of Morgan's Canon derived mainly from its being used to justify the extended reign of Skinnerian behaviorism in animal psychology, many years after the cognitive revolution ousted the behaviorist ruling class in human psychology. If behaviorist interpretations were fair,

Morgan's canon seemed to dictate that they always prevail. Nevertheless, modified versions of the Canon also found support from epistemically careful researchers explicitly interested in the unobservable mental states and cognitive capacities of nonhuman animals (Cheney & Seyfarth 1992; Shettleworth 2010).

In the 21st century, 'anthropomorphism' is still "almost a dirty word in the scientific study of animal cognition" (Shettleworth 2007: 4)[207], even though more researchers have become comfortable attributing psychological capacities to the nonhuman animals they study. Meanwhile, Morgan's Canon has received increased scrutiny from both philosophers and comparative psychologists, many of whom have begun self-identifying as cognitive scientists—'cognitive ethologists' (Allen & Bekoff 1997)—as opposed to behaviorists. Many commenters have pointed out that Morgan's talk of 'higher' and 'lower' psychological faculties relies on a misunderstanding of evolution. This objection is fair enough, but more interesting objections concern (not the wording of but) the principle behind Morgan's Canon, as standardly interpreted. For example, Shettleworth writes that "evolutionary continuity justifies anthropomorphism as a source of hypotheses" (2007: 4), citing Elliot Sober's warning that "if nonhuman animals really are like us in certain respects, the canon may lead us to miss this fact about nature" (1998: 229). Sober worries that being overly cautious about anthropomorphism will lead researchers to an unwarranted form of what Andrews & Huss (2013) have termed 'anthropectomy': the denial that nonhuman animals have

[207] Shettleworth cites Mitchell (2005) and Wynne (2007) to back up this assertion.

human-like psychological characteristics. Jerry Fodor has echoed Sober's worry that Morgan's Canon avoids anthropomorphic bias by falling prey to anthropectic bias: "Why doesn't Fodor's Pop Gun tip the scales equally in the opposite direction? To wit: in no case may we interpret an action as the outcome of the exercise of a lower psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands higher in the psychological scale" (1999: 12).[208]

In assessing the role Morgan's Canon should play in animal cognition research, all of these writers consider the way in which the Canon is currently invoked. Few of them consider the use for which Morgan originally intended his Canon. I think this is a mistake; I am a Canon originalist, insofar as I think Morgan diagnosed the proper role for the attribution of human-like psychological states in animal research back in 1892. The main problem with Morgan's Canon concerns its standard interpretation, not its original meaning. If Morgan's Canon is "perhaps, the most quoted statement in the history of comparative psychology" (Dewsbury 1984: 1987), it is also, as Roger Thomas has quipped, "perhaps the most misrepresented statement in the history of comparative psychology" (Thomas 1998: 156). As Thomas and other historians of science have argued, Morgan himself never intended his canon as a prohibition against anthropomorphism. Instead, he intended it as an empiricist check against bias, given the inevitability of anthropomorphism in any comparative psychology worth its salt (Costall

---

[208] To foreshadow my argument later in this section, note that Fodor's formulations of Morgan's Canon and Fodor's Pop Gun (to be fair, following the wording in Morgan 1894) leave out the word 'fairly'.

1993).

Morgan's writings demonstrate a clear development in their author's thinking about the possibility of animal cognition research. Early in his career, Morgan was an avid reader of George Berkeley (Morgan 1930a), and struck by the power of the epistemological problem of other minds (Morgan 1880), Morgan wrote that

> The results of comparative psychology—the science which has for its object the comparative study of those distorted images of our own mental processes—are incapable of verification … 'Is there a science of comparative psychology?' [I submit] an emphatic negative. (Morgan 1884: 371)

However, Morgan changed his mind within a decade. By the time Morgan devised his Canon, he proudly identified as a comparative psychologist, though this about-face did not keep him from worrying about the epistemological and methodological issues pervading his field. Morgan retained a beef with the way most Darwinists were studying animal minds in the 19th century. Nevertheless, his mature beef was not with anthropomorphism; it was with the practice of inferring mental capacities from anecdotes. He insisted that "observation on one occasion only, no matter how careful and exact that interpretation may be, does not suffice for the interpretation of this or that instance of animal behavior" (Morgan 1930b; excerpted by Cook 1999). The problem identified in this passage is not the attribution of *psychological* capacities to animals, but

the attribution of *any* capacities on the basis of too little evidence. Morgan was obsessed, first and foremost, with empirical rigor.[209]

Empirical rigor is difficult to achieve in comparative psychology, and it was particularly difficult given the way in which Morgan conceived his field of inquiry. Morgan took comparative psychology—the study of animal minds—to necessarily involve the study of the distorted images of human mental processes. According to Morgan's introspectionist ideology, "the first duty of a psychologist is to attain accurate and systematic acquaintance with the working of his own mind, as the cipher in terms of which all other minds must be read" (1894: Chapter 3). An introspectionist comparative psychologist becomes equipped to start studying animal minds only once she has properly understood her own mind via introspection. She then attributes psychological capacities to animals by judging whether "ejections" of her subjective mental processes would make sense of the animals' behaviors (Morgan: 1891). For example, Morgan's psychologist might reason as follows when confronted with my cats around dinner time. 'The cats are meowing crankily, and pacing around their food bowls. Introspection tells me that when I engage in analogous behaviors—complaining, and rooting around in the fridge—it is because I am hungry. Thus, the cats must be hungry.'

Morgan explicitly developed his Canon as subservient to the first duty of the

---

[209] For an amusing (if disturbing) illustration of this obsession, I highly recommend Morgan's (1883a, 1883b, 1887) reports in *Nature* of experiments testing the hypothesis that scorpions commit suicide (which Morgan and other late 19th century naturalists took to directly challenge the theory of evolution by natural selection—if the overarching telos of life is reproduction and survival, then how could an animal have evolved to kill itself?)

introspective method. The Canon says that "in no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be fairly interpreted as the outcome of the exercise of one which stands lower in the psychological scale." In context, this dictum urges nothing more than that researchers rule out alternative explanations. After all, the introspectionist methodology entails that the first explanation Morgan's psychologist devises is always the anthropomorphic hypothesis (that, e.g., the cats are hungry). So, in effect, Morgan's Canon advises the following. 'Do not immediately accept the first explanation you devise via introspection. Before settling on the first explanation you devise, make sure there are not other equally reasonable explanations available.' Maybe the cats see a bug flying around the food bowls. As Morgan stressed, this rule—to always consider reasonable anthropectic hypotheses—is not a brief against anthropomorphism; it is just good empiricism in the face of the anthropomorphic default hypotheses delivered by introspectionism.

Unlike Morgan, I do not think introspection is a reliable—much less necessary—first step in the methodology of animal cognition research. Nevertheless, cognitivist comparative psychologists regularly begin their inquiry by considering the attribution of psychological capacities to animals (though not necessarily psychological capacities shared by the psychologist herself). They are right to do so: animal cognition research must assume it has an object of inquiry. Morgan was correct that inevitably anthropomorphic default hypotheses will always confront the study of animal minds with looming confounds due to anthropomorphic bias. Even in the 21st century, a

suitably modified version of Morgan's Canon serves as a useful corrective to this bias.

Contra Sober and Fodor, there is an asymmetry between anthropomorphic bias and anthropectic bias, since humans are systematically predisposed to the former. Anecdotally, the evidence that humans are inclined towards indulging in anthropomorphism is overwhelming. Everyone with a dog will tell you, at least in their less reflective moments, that Rover wants the biscuit, enjoys playing fetch, despises taking baths, or tries to deceive others by burying his bone. Consider the urge to attribute mental states to the gliders that emerge in Conway's Game of Life.[210] When faced with living animals rather than mere pixels, this urge intensifies. Experimental evidence abets these anecdotes (Heider & Simmell 1944; Kelemen 1999, 2003; Johnson et al. 2001; Casler & Kelemen 2008; Rosset 2008; Kelemen & Rosset 2009). Most strikingly, Kelemen et al. (2013) have demonstrated that "although extended education appears to produce an overall reduction in inaccurate teleological explanation, specialization as a scientist does not, in itself, additionally ameliorate scientifically inaccurate purpose-based theories about the natural world" (2013: 1074). By default, humans, including human scientists, are predisposed to judge animal behavior to be intentional, until presented with evidence to the contrary. Animal cognition researchers ought to take whatever methodological precautions are necessary to counteract this bias towards the anthropomorphic hypotheses that frame their investigations.

So, ought anthropomorphism be *prohibited* in animal cognition research? Should

---

[210] Conway's Game of Life is discussed in Chapter 1.3.2.

researchers *always* accept anthropectic interpretations when available? Of course not. But an openness to appropriate anthropomorphism does not entail an indictment of Morgan's Canon. In its rightful context, Morgan's Canon does not say 'if an action can be fairly interpreted as the outcome of non-anthropomorphic process, accept the anthropectic interpretation'. It says to refrain from accepting the anthropomorphic interpretation when there are other reasonable interpretations on the table.

Consider famous cases of sphexishness (Dennett 1984), wherein animal behaviors appear flexible and intelligent but are actually rote and thoughtless. Many species of birds (most notably plovers) appear to purposively lure predators away from their nests by feigning a broken wing. In these scenarios, the predator, seeing that the bird appears wounded, chases the mother bird instead of finding its nest and eating its eggs or young. When the bird has lured the predator far enough away from the nest, she takes flight, showing the broken wing to be a hoax. This phenomenon is ripe for anthropomorphism: to the casual observer, it seems obvious that the bird is cleverly deceiving the predator in order to protect its hatchlings. However, close observation reveals that plovers automatically engage in broken wing display behaviors whenever predators get close to their nests, whether or not eggs or young are present. Researchers have rightly concluded that broken wing display behaviors are not the product of sophisticated, flexible, and intentional mental processes after all. Properly interpreted, Morgan's Canon does nothing more than direct researchers to engage in precisely this kind of careful observation before jumping to anthropomorphic conclusions.

2.2. Anthropomorphism and attitudes of belief

Cries of THAT'S ANTHROPOMORPHISM! often ring out when people attribute beliefs to non-human animals. I want to suggest that, in most contexts, this accusation gets its fangs only because accusers conflate attitudes of belief with cognitive states of belief. Morgan's Canon is properly aimed at hypotheses about the psychological processes that causally produce behavior. When attempting to understand why some animal behaved in some way, cognitive states—that is, psychological producing causes—often spring to investigators' minds. Why did the cats meow and pace? Why did PAC-MAN pursue the apple? Because they felt hungry. Morgan's Canon helpfully says: 'hold up, animal cognition researcher; best make sure there is not another inner cause that might reasonably be posited as producing those behaviors'.

Meanwhile, when I attribute the belief that it is dinner time to my cats, the skeptic might unhelpfully say 'hold up, Devin; best make sure there is not another cause that might reasonably be posited as producing those behaviors'. The disanalogy between my (unproblematic) belief attribution and the animal cognition researcher's (problematic) belief attribution stems from the fact that I am not positing my cats' beliefs as producing their dinnertime behaviors. Instead, I am attributing an attitude of belief—a pattern of dispositions—to my cats, in order to better understand them as fellow inhabitants of my social environment.[211]

[211] This is not to say that I cannot be wrong! The cats might be pacing because they see a fly, and lack the belief that it is dinner time.

The skeptic about my attribution of beliefs to cats rightly assumes (a) that she knows that humans have attitudes of belief, and (b) that she has a good epistemic handle on those attitudes. She also rightly assumes (c) that it is an open question whether or not nonhuman animals have cognitive states of belief. So far, the skeptic and I have no disagreement. But the skeptic goes on to wrongly assume (d) that attitudes of belief can be safely conflated with cognitive states of belief. Combining this bad assumption (d) with her good assumptions (a–c), she makes a few inferences to dubious conclusions. She infers that humans have cognitive states of belief, and that she has a good epistemic handle on these cognitive states.[212] She also infers that it is an open question whether or not nonhuman animals have attitudes of belief.

By rejecting the conflation of attitudes of belief with cognitive states of belief, Ryleanism provides a metric for the folk practice of attributing beliefs to nonhuman animals, without making any unsubstantiated empirical assumptions about their cognitive capacities. Morgan's Canon—a dictum guiding the cognitive scientific attribution of producing causes—does not warn directly against attributions of attributor-relative attitudes.

Morgan's Canon *does* warn, *indirectly*, against *incautious* attributions of attitudes of belief. After all, there is always the danger of illicitly anthropomorphizing the *dispositions* that make up attitudes. For example, I will wrongly attribute belief to my

[212] To put a face to this skeptic, consider Fodor's Granny (1987: 6), who he swears has forgotten more about belief than any philosopher will ever know.

thermometer if I interpret it as being disposed (among other things) to enthusiastically report the temperature with the goal of informing Devin how cold it is. That is where the proper use of Morgan's Canon comes in handy. We should be wary of anthropomorphic interpretations of behavior when anthropectic interpretations are on the table; as a matter of fact, the thermometer's report of the temperature does not resemble a human's report of the temperature at all. An anthropectic interpretation (or at least an interpretation that invokes the intentions of the designer of the thermometer rather than the thermometer itself) is much more reasonable. Similarly, I *may* wrongly ascribe a piping plover the belief that feigning a broken wing distracts predators, if I interpret the plover as being disposed (among other things) to purposively feign a broken wing because it occurrently judges this trickery to be the ideal way to lure a predator away from its babies. Even if the plover does have the attitude of belief in question, it certainly does not live out the belief according to my mentalistic model thereof. Properly interpreted, Morgan's Canon helps us avoid such facile anthropomorphic understandings of animal behavior, and thus indirectly informs lay belief attribution.

3. Brute believers

Do nonhuman animals believe?

There are tricky issues concerning whether or not nonhuman animals have cognitive states of belief. Even if cognitive scientists agreed about whether or not humans have cognitive states of belief, they would not agree about the nature of those

cognitive states. Do cognitive states of belief necessarily involve language, or higher order thought (Davidson 1982)? Full-blown concepts (Stich 1983)? The ability to subconsciously perform predicate logic (Bermudez 2003)? If so, certain animals might have proto-beliefs (per Bermudez) as opposed to full-blown human-like cognitive states of belief, or they may lack cognitive states of belief altogether (per Davidson). Addressing these issues lies beyond our current purview.

The question within our purview—of whether animals have attitudes of belief—is less problematic. Ryleanism does justice to the (common) folk intuition that many animals boast attitudes of belief, without simply discarding the (not uncommon) intuition that they do not. Consider these intuitions in turn.

First, many people would find it absurd to deny that my cats believe it is dinner time. Consider Raimond Gaita's appeal to this kind of intuition.

> I don't *conjecture* whether [a dog] is the kind of creature who is sometimes warm and sometimes cold, who sometimes has pleasures and is sometimes in pain, who sometimes believes one thing and hopes or fears another. Nor do I *assume* it, or *take it as certain*—that is, for practical though not for philosophical or scientific purposes. I am *absolutely certain*; that is to say, I have not the slightest doubt. (Gaita 2002: 44)

We do not know what cognitive mechanisms underlie cats and dogs' attitudes of belief;

nor do we know whether the same cognitive mechanisms underlie human and nonhuman beliefs. Nevertheless, it seems to be an affront to quotidian experience to deny that cats and dogs take the world to be certain ways and not others.

This denial would also involve ignoring the boatloads of empirical evidence showing that many nonhumans are disposed to represent their environments in flexible and goal-directed manners (Saidel 2009). Ryleanism makes good on this intuition—and the evidence backing it up—by allowing that animals of different species can share the same belief even in the face of considerable differences in cognitive architecture. So long as different animals live in patterns that sufficiently fit a belief attributor's general-purpose stereotype, they believe the same thing. Nonhuman animals do not believe in a derivative manner. As Gaita argues, "our ways of speaking about knowledge and belief have not been first and fully formed just in our lives with human beings and then applied conjecturally with animals" (2002: 72). Instead, our general-purpose models of particular beliefs are constructed and refined to cross (some) species boundaries, especially when we grow up with animals around.

At the same time, belief attributors construct individualized models of belief for different types of believers.[213] Unlike non-attributor-relative accounts of belief, Ryleanism leaves room for animals of different species to believe in different styles. For example, my model believing-it-is-dinner-time-like-a-cat differs from my model of

---

[213] Models of assent (see Chapters 2.2.6 and 6.3.4), for instance, are a particular variety of specialized model of belief which cast humans as appropriately specialized believers (that is, assenters) and plausible exclude all nonhuman animals.

believing-it-is-dinner-time-like-a-human, and this outer recognition of styles of belief allows me to understand my cats more subtly than if I had solely wielded the blunt tool that is my general-purpose stereotype for cross-species-belief-that-it-is-dinner-time. Ryleanism thus provides flexible tools for accounting for the subtleties of human practices of belief attribution to nonhuman animals.

Ryleanism also elucidates why nagging intuitive skepticism that nonhuman animals are believers is justified. Some behaviors characteristic of believing differ dramatically from species to species. As such, members of some species (lobsters, maybe) might sufficiently fit intraspecific models of belief but not interspecific models, and members of other species (coral, maybe) might not sufficiently fit any models of belief. The question of animal belief must be taken up on a case by case basis, informed by both the propensities of candidate believers and the models wielded by belief attributors. Further complicating matters, there is no solution to the epistemological problem of other minds. Belief attributors do not have access to the thoughts and feelings of other animals; we can never know for sure how well animals fit our models of belief, because we can never know with absolute certainty that they possess the relevant phenomenological dispositions. Moreover, in the context of belief attribution to nonhuman animals, the problem of other minds is not merely a radical skeptical hypothesis. Putting radical skepticism aside, it should be uncontroversial that dogs and cats feel pain. But even while putting radical skepticism aside, the problem of other minds engenders the frank assessment that we cannot know—at least given the current

state of the evidence in cognitive psychology and neuroscience—whether dogs are disposed to consciously isolate abstract qualities, or cats are disposed to entertain representational thoughts. Without this knowledge, there will always be lingering epistemological questions about our attributions of belief to nonhuman animals (as well as our attributions of belief to humans lacking linguistic and other cognitive faculties).

4. Monkey mindreaders

Are nonhuman animals belief attributors?

Psychologists trying to figure out who qualifies as a full-blown belief attributor usually rely on false belief tasks, though several researchers have plumped for more variety in benchmarks (Gopnik et al. 1991; Boesch 2007; Andrews 2017). Until recently, some intriguing results notwithstanding (Crockford et al 2012), the tenuous consensus has been that nonhuman animals are uniformly cognitively unequipped to pass the false belief task. Enter Krupenye, Kano, Hirata, Call, and Tomasello (2016).

Inspired by work on human infants (Southgate et al 2007), Krupenye, Kano, et al. introduced eye-tracking as a method of testing for nonhuman false belief attribution. Chimpanzees, bonobos and orangutans have the ability to pass two different false belief tasks, thereby (purportedly) demonstrating that humans are not the only great ape belief attributors.[214]

[214] For years, Michael Tomasello's lab had made Morgan proud, producing empirical results in support of the conclusion that humans are unique in their ability to attribute beliefs. This

Ape subjects in both trials watched movies in which agents formed false beliefs about the location of objects. Using an infrared eye-tracker, investigators recorded the parts of the movies to which the apes paid visual attention. These recordings reveal that the apes' gaze regularly anticipated that the agents would look for the objects where they had last seen them, rather than where they actually were.[215] Krupenye, Kano, et al. write that these findings "suggest that apes solved the task by ascribing a false belief to the actor, challenging the view that the ability to attribute reality-incongruent mental states is specific to humans" (113). The researchers claim that the apes' anticipatory gaze is best explained by the hypothesis that they use the attribution of belief to successfully predict behavior.

Cecilia Heyes has objected that the findings of Krupenye, Kano, et al. are equally compatible with the hypothesis that apes are (not belief attributors but) mere 'submentalizers': creatures with the ability to predict behavior by way of low-level, domain-general psychological processes. On Heyes's interpretation, apes visually anticipate where agents will look for objects by picking up on behavioral cues that help them track agents' dispositions, without ever explicitly attributing the beliefs that tie those dispositions together. For example, Heyes suggests that the apes in the Krupenye, Kano, et al. study may have tracked "the appearance and disappearance of the striking green shirt" (2017: 2) worn by the agent, and associated the reappearance of the green

---

anthropectic history made their recent announcement of anthropomorphic results all the more compelling.

[215] Videos available at http://science.sciencemag.org/content/suppl/2016/10/06/354.6308.110.DC1.

shirt with the location of the object when the green shirt was last on the scene.[216] In reply, Krupenye, Kano, et al. (2017) have followed Heyes's methodological suggestions and replicated their 2016 study while controlling for some features of the submentalizing hypothesis.

Although they disagree about whether apes are belief attributors, Heyes and Krupenye, Kano, et al. agree that this disagreement cuts right to the heart of the bigger question of how (indeed, whether) apes understand other minds. Tomasello has remarked that being able to pass the false belief task "means understanding that there exists a mental world distinct from the physical world" (quoted in *Duke Today* 2016). According to this mainstream line of thinking, if apes attribute beliefs, then they grasp that there is something going on inside other minds. If apes merely submentalize, then they track dispositions associated with other minds, but do not recognize inner states themselves. Properly controlled false belief tasks are supposed to provide evidence to pull these dichotomous possibilities apart.

However, several outspoken researchers are skeptical about the explanatory power of false belief tasks. According to these skeptics, there is a "logical problem" (Povinelli & Vonk 2003; Hurley & Nudds 2006; Penn & Povinelli 2007) with the methodology of false belief tasks. As Robert Lurz (2011) argues, most false belief tasks systematically fail to distinguish the hypothesis that animals attribute beliefs from the

[216] Andrews (2017), invoking Heider and Simmel (1944), argues that the methodology suggested by Heyes results from flawed theorizing about apes' understanding of agency.

hypothesis that they attribute complex behavioral dispositions, and thus merely submentalize. According to skeptics, this failure is not a mere methodological problem to be solved by further experimental controls; it is an experimentally intractable problem that arises from the epistemic logic of false belief tasks. By running false belief tasks, experimenters are attempting to infer belief attribution from the observable behaviors of candidate belief attributors. But, logically speaking, all possible observable behaviors could be explained by the alternative hypothesis that subjects are mere submentalizers as opposed to full-blown belief attributors. Thus, false belief tasks, including those used by Krupenye, Kano, et al. (2016), can never settle their target question—or even more bleakly, if researchers running false belief tasks follow the letter of Morgan's Canon, then they must always deny the capacity to attribute beliefs to animals.

However, just as charges of anthropomorphism in belief attribution get their fangs from the conflation of attitudes and cognitive states, the logical problem gets its fangs from the paramechanical understanding of attitudes of belief as unobservable causes productively intervening between observable stimuli and behaviors. If beliefs are inner causes, then false belief tasks will never enable researchers to distinguish belief attribution from submentalizing that solely tracks observable properties.[217]

My version of Ryleanism dissolves the logical problem by rejecting Davidson's dogma (according to which belief attributors must construe beliefs as inner causes). I also reject any conception of submentalizing that precludes the tracking of inner causes.

---

[217] Nor will false belief tasks' more sophisticated cousins do the trick (Buckner 2014).

In a Rylean framework, the logical problem is ill-posed. Attitudes of belief are identical to patterns of dispositions to act, react, think, and feel. To submentalize is to track some of the behavioral, cognitive, and phenomenal dispositions that might partly compose beliefs. To attribute a belief is to recognize a pattern among those dispositions. Heyes's question about whether the apes in the Krupenye, Kano, et al. study are attributing beliefs or merely submentalizing is a good one, but it should not be cast as the question of whether they are attributing inner producing causes or outer behavioral dispositions. That question would, indeed, be susceptible to the logical problem, just as nearly all questions about inner states are susceptible to the problem of other minds. Instead, the relevant question is whether the apes recognize whole patterns of dispositions as states of taking the world to be some way, or pick up on individual dispositions in a less systematic manner. In other words, are the apes wielding models of beliefs? Or are they merely tracking and responding to dispositions that would go into models of beliefs? I suspect that there are empirically tractable differences between these hypotheses, having to do with the generality and exportability of the apes' social practices. For example, if apes were able to perform similarly on a variety of eye-tracking false belief tasks—tasks that systematically varied the observable characteristics of the situation while requiring successful subjects to track the same belief—then this would provide evidence that they were attributing beliefs rather than tracking individual dispositions.[218]

---

[218] Supplementing the practical suggestions put forward by Heyes (2015), Apperly & Butterfill (2013) provide an excellent discussion of how to distinguish experimentally between belief attribution and mere submentalizing (or, as they call it, 'minimal mindreading').

Whether animals attribute inner thoughts (such as tokens of inner speech) or feelings (such as aches and pangs) to each other is a fascinating question. They might do so in a systematic fashion, weaving these inner states into models of their conspecifics' attitudes and traits, or they might do so by submentalizing, honing in on a specific conspecific's feeling of pain as she howls. Alternatively, nonhuman animals might be ignorant of the existence of phenomenologies other than their own. Fascinating though it is, this is not a question that researchers need to answer in order to make progress on the animal mindreading debate. Studies like Krupenye, Kano, et al. (2016) and Crockford et al. (2012) have provided preliminary evidence that nonhuman great apes attribute beliefs to one another, whether or not they attribute inner states.

However, researchers should not be too quick to jump to conclusions. Heyes might be right that great apes do not use belief attribution in isolated Krupenye-Kano type situations; submentalizing may be equal to any given false-belief task. Moreover, as Kristin Andrews (2017) has pointed out, there is good evidence (Hare et al. 2000; Karg et al. 2015) that great apes attribute visual access to others.[219] Whereas believing is a general state of taking the world to be some way, not tied to any particular sense modality, seeing is the limited state of taking the world to be some way through occurrent vision. Whether seeing is (sufficient for) believing depends on the model of belief at play. Importantly, an animal might have a model of seeing without having any models of believing whatsoever. The apes in the Krupenye, Kano, et al. study might

[219] Bugnyar et al. (2016) provide evidence that crows attribute states of seeing too.

have accomplished their task (not by attributing beliefs or submentalizing but) by attributing visual access—and the memory of visual access—to the agent. To make real strides on the question of whether or not nonhuman animals attribute beliefs, researchers will have to figure out which situations demand that the animals under investigation engage in belief attribution, as opposed to either submentalizing or attributing mental states other than beliefs.

Moreover, basing comparative psychological inquiry on human models of attitudes of belief can only get researchers so far. Belief—the general (as opposed to sense-modality-specific) state of taking the world to be some way—is a fairly stable category via which humans, across cultures, understand each other. However, if relativistic Ryleanism makes any contributions to animal cognition research, foremost among them ought to be the revelation that different belief attributors—particularly belief attributors who differ enough to belong to different species—are likely to construe beliefs in different ways, and thereby metaphysically constitute beliefs in different ways. Rylean research on nonhuman animal mindreading would focus on investigating how particular nonhuman animals understand each other, rather than narrowly focusing on whether, like humans, any nonhuman animals understand each other in terms of human-like models of belief. Even if nonhuman animals are belief attributors, their models might differ from humans' in surprising ways. If, rather than pursuing the Human Uniqueness Project, researchers are interested in understanding the psychological capacities of nonhuman animals in their own right, then they should be

interested in figuring out the idiosyncratic contours of the ways in which particular (species of) animals understand other minded agents. There is only so much that can be gleaned by way of an exclusive focus on figuring out how closely nonhumans' social cognitive practices resemble our own.

5. Anthropomorphism redux

I am far from against anthropomorphism. As Morgan recognized, thinking about nonhuman animal minds in terms of human attitudes and cognitive processes is often the only way to get animal psychology off the ground. Nevertheless, I am against anthropocentrism in comparative psychology.[220] Any research program that takes the question 'what makes humans unique within the animal kingdom?' as its alpha and omega will obscure as much about nonhuman animal cognition as it illuminates. I am interested in octopus social practices—such as they are—qua octopus social practices, not qua potentially proto-human social practices.

I am thus also against any presumption in favor of anthropomorphism concerning nonhuman belief attribution. Due consideration of Morgan's Canon (shed of its original introspectionist trappings) should lead us to consider not only behavioristic interpretations of the evidence about animal behavior, but also non-anthropomorphic mentalistic interpretations. Perhaps some nonhuman animals are belief attributors, just like humans. Before settling for that default hypothesis, however, we should use

[220] Buckner (2013) provides a useful discussion of three varieties of anthropocentrism.

Morgan's Canon, against our default anthropomorphism, to explore alternative, nondoxastic ways in which nonhuman animals might model each others' minds. To pit belief attribution against behavior attribution is to support a false and counterproductive dichotomy. Some nonhuman animals are mindreaders—attitude attributors—even if they are not belief attributors, and perhaps even if they do not attribute any inner states to other animals. For example, many animals attribute states of seeing to their conspecifics—including predominantly asocial animals which are unlikely to be belief attributors, like octopi (Godfrey-Smith 2016: 71).

If nonhuman animals do not attribute beliefs to one another, then this is not necessarily because they are cognitively deficient (relative to human beings). Instead, it might just be that different animals model minds differently, to suit their different needs in their different social environments. Even if (as I suspect) some nonhuman animals do attribute beliefs to each other, the details of their models of beliefs doubtlessly differ in some drastic respects from species-typical human models. As Morgan aptly wrote, "given two different minds and the same facts, how different are the products!" (1891: 335). For example, I am confident in speculating that nonhuman animals do not, as a rule, take beliefs to involve dispositions to assent to propositions, much less to involve dispositions to assent to propositions linguistically. Philosophers of animal minds and comparative psychologists would do well to heed the differences in how different animals construe each others' minds, without antecedently setting human belief attribution as the interspecific standard.

6. Conclusion: minding our differences

That is as good a point as any with which to conclude. As lay people and social animals, our interest in each others' minds is, in large part, an interest in difference. One moral of the last 379 pages is that the same should be true when we wear our cognitive scientist and philosopher hats. Partly because minds are hodgepodges, partly because of our intrinsic cognitive differences, and partly because aspects of our minds—including our attitudes of belief—are constructed in relation to our own and others' models of minds, the details of mental differences will inevitably pervade any metaphysics of mind that aspires to completion. As argued in Part I, metaphysicians of mind (as well as psychologists) have often erred in conflating attitudes with cognitive states, rather than recognizing that the general umbrella of 'the mental' covers various incommensurable categories of states, processes, and experiences. That attitudes of belief ought not be conflated with cognitive states of belief is perhaps most vividly illustrated by the potentially enormous gulf between how different animals understand each other as minded. As argued in Chapter 6, attributor-relative attitudes (including beliefs) vary in style between believers. And as argued in Chapters 7 and 8, attitudes vary between interpreters who model attitudes differently. The philosophical task of understanding the mind, and thereby addressing the substance of the mind-body problem, does require attention to the generic, universal mark(s) of the mental. I hope to have shown that it also benefits from attention to the provenance and extent of various kinds of mental diversity.

Bibliography

Adams, R. (2006). *A Theory of Virtue: Excellence in Being for the Good*. OUP.

Allen, C. & Bekoff, M. (1997). *Species of Mind: The Philosophy and Biology of Cognitive Ethology*. MIT Press.

Alfano, M. (2013). *Character as Moral Fiction*. CUP.

Ames, D.R. (2004). "Inside the mind reader's tool kit: projection and stereotyping in mental state inference", *J Pers. Soc. Psychol,* 87(3): 340–353.

Amodio, D.M. (2014). "The neuroscience of prejudice and stereotyping", *Nature Reviews Neuroscience*, 15:10, 670–682.

Andrews, K. (2012). *Do Apes Read Minds?* MIT Press.

Andrews, K. (2015). "The Folk Psychological Spiral: Explanation, Regulation, and Language", *Southern Journal of Philosophy*, Volume 53, S1, 50–67.

Andrews, K. (2017). "Chimpanzee mindreading: Don't stop believing", *Philosophy Compass,* 12(1).

Andrews, K. (2018). "Apes track false beliefs but might not understand them", *Learning & Behavior*, 46: 1, 3–4.

Andrews, K & Huss, B. (2013). "Anthropomorphism, Anthropectomy, and the Null Hypothesis." *Biology and Philosophy*, 29(5), 711–729.

Anscombe, E. (1957). *Intention.* HUP.

Apperly, I. (2011). *Mindreaders*. Psychology Press.

Apperly, I. & Butterfill, S. (2009). "Do humans have two systems to track beliefs and belief–like states?" *Psychological Review* 116(4): 953–70.

Appiah, K.A. (1985). "The Uncompleted Argument: Du Bois and the Illusion of Race" *Critical Inquiry*, Vol. 12, No. 1, 21–37.

Antony, L. (2013). "'Human Nature' and its Role in Feminist Theory" in *Arguing about Human Nature,* (Eds. Downes & Machery), Routledge, 521–542.

Aristotle (350 B.C.E.). *De Anima*.

Armstrong, D. (1968). *A Materialistic Theory of the Mind*. London: RKP.

Audi, R. (1972). "The concept of believing", *Personalist*, 53, 43–62.

Audi, R. (1994). "Dispositional beliefs and dispositions to believe", *Noûs*, 28, 419–434.

Austen, J. (1813). *Pride & Prejudice*. Dover.

Avramides, A. (2002). *Other Minds.* Routledge.

Baier, A. (1985). *Postures of the mind*. Minneapolis: University of Minnesota.

Baker, L.R. (1987). *Saving Belief: A Critique of Physicalism*. PUP.

Baker, L.R. (1995). *Explaining Attitudes.* MIT Press.

Baker, L.R. (2003). "Belief Ascription and the Illusion of Depth", *Facta Philosophica* 5, 183–201.

Baker, L.R. (2013). *Naturalism and the First–person Perspective.* OUP.

Baron–Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press.

Bartsch, K. & Wellman, H. (1995). *Children Talk About the Mind*. OUP.

Bermúdez, J. (2003). *Thinking Without Words*. OUP.

Birch, S. & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychol. Sci.*; 18(5): 382–6.

Block, N. (1990). "Inverted Earth," in Tomberlin, ed., *Philosophical Perspectives*, Atascadero, CA: Ridgeview Press.

Block, N. & Fodor, J. (1972). "What Psychological States Are Not", *Philosophical Review*, 81: 159– 181.

Boesch, C. & Boesch–Achermann, H. (2000). *The Chimpanzees of the Taï Forest: Behavioural Ecology and Evolution.* OUP.

Boesch, C. (2009). *Wild Cultures: A Comparison between Chimpanzee and Human Cultures*. CUP.

Bogdan, R. (1997). *Interpreting Minds*. MIT Press.

Bogdan, R. (2000). *Minding Minds*. MIT Press.

Bogdan, R. (2008). *Predicative Minds*. MIT Press.

Boghossian, P. and Velleman, D. (1989). "Colour as a secondary quality", *Mind* 98:81–103.

Boghossian, P. (2006). "What is Relativism?" in P. Greenough and M. Lynch (eds.), *Truth and Realism*. OUP.

Boghossian, P. (2007) *Fear of Knowledge: Against Relativism and Constructivism*. OUP.

Bohl, V. (2015). "We read minds to shape relationships", *Philosophical Psychology*, 28:5 674–694.

Boudry, M. & Coyne, J. (2016). "Disbelief in belief: On the cognitive status of supernatural beliefs", *Philosophical Psychology 29*, 601–615.

Buckner, C. (2013). "Morgan's Canon, Meet Hume's Dictum: Avoiding Anthropofabulation in Cross–Species Comparisons" *Biology & Philosophy* 28:5, 853–871.

Buckner, C. (2014). "The Semantic Problem(s) with Research on Animal Mindreading" 29:5, 566–589.

Bugnyar, T., Rebar, S., & Buckner, C. (2016). "Ravens Attribute Visual Access to Unseen Competitors", *Nature Communications* (7).

Burge, T. (2010). *Origins of Objectivity*. OUP.

Buss, A.R. (1978). "Causes and reasons in attribution theory: a conceptual critique", *Journal of Personality and Social Psychology* 36:1311–1321.

Buttelmann, D., M. Daum, N. Zmyj, & M. Carpenter (2013). "Selective Imitation of In–Group over Out–Group Members in 14–Month–Old Infants", *Child Development*, Vol 84: No 2, 422–428.

Butterfill, S. & Apperly, I. (2013a). "How to Construct a Minimal Theory of Mind" *Mind and Language* 28 (5):606–637.

Butterfill, S. & Apperly, I. (2013b). "Replies to Three Commentaries on Minimal Theory of Mind", *Brains Blog*, http://philosophyofbrains.com/wp–content/uploads/2013/11/replies–to–commentaries–on–minimal–theory–of–mind.pdf

Byrne, A. (1998). "Interpretivism", *European Review of Philosophy* 3.

Callard, A. (2018). *Aspiration: The Agency of Becoming*. OUP.

Call, J. & Tomasello, M. (2008). "Does the chimpanzee have a theory of mind? 30 years later", *Trends in Cognitive Science,* 12(5):187–192.

Camp, E. (2015). "Logical Concepts and Associative Characterizations", in E. Margolis & S. Laurence (eds.), *The Conceptual Mind: New Directions in the Study of Concepts,* MIT Press, 592–621.

Camp, E. (forthcoming). "Imaginative Frames for Scientific Inquiry: Metaphors, Telling Facts, and Just–So Stories", in P. Godfrey–Smith & A. Levy (eds.), *The Scientific Imagination*. OUP.

Carruthers, P. (2006). *The architecture of the mind.* OUP.

Carruthers, P. (2015a). *The Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*. OUP.

Carruthers, P. (2015b). "Two Systems for Mindreading?" *Review of Philosophy and Psychology*, 7:141–162.

Casler, K. & Kelemen, D. (2008). Developmental continuity in the teleo–functional bias: Reasoning about nature among Romanian Roma adults (Gypsies). *Journal of Cognition and Development*, *9*, 340–362.

Catechism of the Catholic Church. (1992). http://www.vatican.va/archive/eng0015/_index.htm

Chase, P. (2013). “Human Cultures is more than Memes in Transmission” in (Hatfield, G. & Pittman, H. eds.), *Evolution of Mind, Brain and Culture,* Penn, 347–378.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT Press.

Cheney, D. & Seyfarth, R. (1992). *How Monkeys See the World*. Chicago University Press.

Cheney, D. & Seyfarth, R. (2007). *Baboon Metaphysics*. Chicago University Press.

Child, W. (1994). *Causality, Interpretation, and the Mind*. OUP.

Chirimuuta, M. (2015). *Outside Color*. MIT Press.

Choi, S. and Fara, M. (2014). "Dispositions", *The Stanford Encyclopedia of Philosophy*.

Churchland, P.M. (1970). “The Logical Character of Action–Explanations” *The Philosophical Review*, Vol. 79, No. 2, 214–236.

Churchland, P.M. (1979). *Scientific Realism and the Plasticity of Mind*. CUP.

Churchland, P.M. (1981). “Eliminative Materialism and the Propositional Attitudes” *Journal of Philosophy* Vol. 78, No. 2., 67–90.

Churchland, P.M. (1984). *Matter and Consciousness* MIT Press.

Churchland, P.M. (1993). “Evaluating Our Self Conception,” *Mind and Language* 8, 2: 211–222.

Churchland, P.S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. MIT Press.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. OUP.

Cohen, J. (2004). “Color Properties and Color Ascriptions: A Relationalist Manifesto” *The Philosophical Review*, Vol 113 No 4, 451–506.

Cohen, J. (2009). *The Red and the Real.* OUP.

Cohen, L.J. (1992). *Belief and Acceptance*. OUP.

Connolly, K. (2017). "Perceptual Learning", *The Stanford Encyclopedia of Philosophy*.

Costall, A. (1993). How Lloyd Morgan's Canon backfired. *Journal of the History of the Behavioural Sciences,* 29, 113 –122.

Cook, R. (1999). "Morgan's comments on the need to see the process leading to an "intelligent" behavior", http://pigeon.psy.tufts.edu/psych26/morgan.htm

Crane, T. (2016). *The Mechanical Mind*, Third Edition. Routledge.

Crockford, C., Wittig, R., Mundry, R., & K. Zuberbuhler (2012). "Wild Chimpanzees Inform Ignorant Group Members of Danger", *Current Biology* 22:2, 142–146.

Csibra, G. & Gergely, G. (2013). "Teleological Understanding of Actions", *Navigating the Social World,* OUP, ed. Banaji & Gelman, 38–43.

Cummins, R. (1975). "Functional analysis" *Journal of philosophy* 72:741–60.

Cummins, R. (1983). *The Nature of Psychological Explanation*. MIT Press.

Cummins, R. (1996). *Representations, targets, and attitudes*. MIT Press.

Curry, D.S. (2018). "Cartesian Critters Can't Remember", *Studies in History and Philosophy of Science*. Online First.

Darley, J. M., and Batson, C.D. (1973.) "From Jerusalem to Jericho": A study of Situational and Dispositional Variables in Helping Behavior". *JPSP*, 27, 100–108.

Davidson, D. (1963). "Actions, Reasons, and Causes" in *Essays on Actions and Events*, OUP (2001), 3–19.

Davidson, D. (1967). "Causal Relations" *The Journal of Philosophy*, Volume LXIV, 21, 691–703.

Davidson, D. (1968). "On Saying That" in *Inquiries into Truth and Interpretation*, OUP (2001), 93–108.

Davidson, D. (1970). "Mental Events" in *Essays on Actions and Events*, OUP (2001), 207–229.

Davidson, D. (1973). "Radical Interpretation" in *Inquiries into Truth and Interpretation*, OUP (2001), 125–140.

Davidson, D. (1974). "Psychology as Philosophy" in *Essays on Actions and Events*, OUP (2001), 229–244.

Davidson, D. (1975). "Thought and Talk" in *Inquiries into Truth and Interpretation*, OUP (2001), 155–170.

Davidson, D. (1980). "A Unified Theory of Thought, Meaning, and Action" in *Problems of Rationality*, OUP (2004), 151–166.

Davidson, D. (1982). "Rational Animals" *Dialectica* Vol 36:4, 317–327.

Davidson, D. (1983). "A Coherence Theory of Truth and Knowledge" in *Subjective, Intersubjective, Objective*, OUP (2001), 137–153.

Davidson, D. (1989). "The conditions of thought", in J. Brandl & W. Gombocz (Eds.), *The mind of Donald Davidson* (pp. 193–200). Rodopi.

Davidson, D. (1990). "The structure and content of truth", *The Journal of Philosophy*, 87(6), 279–328.

Davidson, D. (1991). "Three Varieties of Knowledge" in *Subjective, Intersubjective, Objective*, OUP (2001), 205–220.

Davidson, D. (1995). "Could there be a science of rationality?" in *Problems of Rationality*, OUP (2004), 117–134.

Davidson, D. (1997). "The Emergence of Thought" in *Subjective, Intersubjective, Objective*, OUP (2001), 123–134.

Davidson, D. (2001a). *Inquiries into Truth and Interpretation*. OUP.

Davidson, D. (2001b). *Subjective, Intersubjective, Objective*. OUP.

Davidson, D. (2001c). *Essays on Actions and Events*. OUP.

Davidson, D. (2004). *Problems of Rationality*, OUP.

Davidson, D. (2005). *Truth, Language and History: Philosophical Essays*. OUP.

Davies, P.S. (2001). *Norms of Nature: Naturalism and the Nature of Functions.* MIT Press.

Davies, P.S. (2009). *Subjects of the World.* Chicago University Press.

Dennett, D.C. (1969). *Content and Consciousness*, Routledge & Kegan Paul.

Dennett, D.C. (1971). "Intentional Systems", *The Journal of Philosophy*, 68:4, 87–106.

Dennett, D.C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. MIT Press.

Dennett, D.C. (1984). *Elbow Room: The Varieties of Free Will Worth* Wanting. MIT Press.

Dennett, D.C. (1987). *The Intentional Stance.* MIT Press.

Dennett, D.C. (1991a). "Real Patterns", *The Journal of Philosophy*, Vol 88, No. 1, 27–51.

Dennett, D.C. (1991b). *Consciousness Explained*. Little, Brown.

Dennett, D.C. (1993). "The message is: there is no medium" (reply to Jackson, Rosenthal, Shoemaker & Tye), *Philosophy & Phenomenological Research*, 53, (4), 889–931.

Dennett, D.C. (1994). "Get Real", *Philosophical Topics*, 22:1&2, 505–568.

Dennett, D.C. (1995). *Darwin's Dangerous Idea*, Simon & Schuster.

Dennett, D.C. (1996). *Kinds of Minds*, Basic Books.

Dennett, D.C. (1998). *Brainchildren*. MIT Press.

Dennett, D.C. (2008a). "Autobiography (Part One)" *Philosophy Now*, issue 68.

Dennett, D.C. (2008b). "Autobiography (Part Two)" *Philosophy Now*, issue 69.

Dennett, D.C. (2009). "Intentional Systems Theory" in Brian McLaughlin (ed.), *Oxford Handbook in the Philosophy of Mind*. OUP.

Dennett, D.C. (2013). *Intuition Pumps and Other Tools for Thinking*. Norton.

Dennett, D.C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. Norton.

de Sousa, R. (1971) "How to Give a Piece of Your Mind: Or, the Logic of Belief and Assent", *Review of Metaphysics* 25 (1):52–79.

Dewey, J. (1925). *Experience and Nature*. Dover.

Dewsbury, D.A. (1984). *Comparative Psychology in the Twentieth Century*. Hutchinson Ross.

Dixon, E. (1982). "The Limits of Animal Intelligence", *Nature* 46:1191, 392–393.

Doris, J. (2002). *Lack of Character.* CUP.

D'Oro, G. (2012). "Reasons and causes: The philosophical battle and the metaphilosophical war", *Australasian Journal of Philosophy* 90, 2, 207–21.

D'Oro, G. & Sandis, C. (2013). *Reasons and Causes: Causalism and Anti–Causalism in the Philosophy of Action*, Palgrave MacMillan, introduction.

Drayson, Z. (2014). "The Personal/Subpersonal Distinction" *Philosophy Compass,* Volume 9, Issue 5, 338–346.

Dretske, F. (1981). *Knowledge and the Flow of Information.* MIT Press.

Dretske, F. (1986). "Misrepresentation", in (R. Bogdan ed.) *Belief: Form, Content, and Function*, OUP, 17–36.

Dretske, F. (1988). *Explaining Behavior.* MIT Press.

Dretske, F. (1995). *Naturalizing the Mind.* MIT Press.

Dretske, F. (1991). "Reply to Baker" *Philosophical Studies,* 63:1, 113–117.

Dretske, F. (2000). *Perception, Knowledge and Belief.* Cambridge Studies in Philosophy.

DukeToday. (2016). "Apes understand that some things are all in your head", https://today.duke.edu/2016/10/falsebeliefs.

Du Bois, W.E.B. (1897). "The conservation of races". American Negro Academy.

Dunn, J. & Brown, J. (1993). "Early conversations about causality: Content, pragmatics and developmental change", *British Journal of Developmental Psychology*, II, 107–123.

Egan, F. (2014). "Explaining representation: a reply to Matthen": *Philosophical Studies* 170:137–142.

Engel, P. (2008). "Davidson on Epistemic Norms" in *Knowledge, Language and Interpretation: On the Philosophy of Donald Davidson*, Amoretti and Vassallo (eds.), Ontos Verlag: 123–146.

Enger, A. (2010). "To Kill a Mockingbird: Fifty Years of Influence on the Legal Profession." In *Harper Lee's To Kill a Mockingbird: New Critical Essays.* Ed. Michael J. Meyer. Lanham, MD and Toronto: Scarecrow Press, 2010:66–80.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). "Perspective taking as egocentric anchoring and adjustment" *Journal of Personality and Social Psychology*, 87, 327–339.

Estes, D., Wellman, H. & Wooley, J. (1989). Children's understanding of mental phenomena. In H. Reese (Ed.), *Advances in child development and behavior, Vol 22.* (41–87), Academic Press.

Evans, J. (2008). "Dual–Processing Accounts of Reasoning, Judgment, and Social Cognition", *Annu. Rev. Psychol.* 59:255–78.

Falvey, K. (1999). "A Natural History of Belief", *Pacific Philosophical Quarterly*, Vol. 80, No. 4, 324–345.

Fara, M. & Choi, S. (2012). "Dispositions", *The Stanford Encyclopedia of Philosophy*.

Feigl, H. (1958). "The 'mental' and the 'physical'", *Minnesota Studies in the Philosophy of Science Volume II: Concepts, Theories, and the Mind–Body Problem*.

Fiske, S. & Taylor, S. (2013). *Social Cognition: From Brains to Culture*. Sage.

Flanagan, O. (1991). *Varieties of Moral Personality*. HUP.

Fodor, J. (1975). *The Language of Thought.* HUP.

Fodor, J. (1980). "Methodological Solipsism Considered as a Research Strategy in Cognitive Science," *Behavioral and Brain Sciences*, 3: 63–73.

Fodor, J. (1983). *The Modularity of Mind.* MIT Press.

Fodor, J. (1987). *Psychosemantics.* MIT Press.

Fodor, J. (1992). *The Elm and the Expert.* MIT Press

Fodor, J. (1999). "Not so Clever Hans", *London Review of Books*, February 4, 12–13.

Fodor, J. & Lepore, E. (1992). *Holism*. OUP.

Fodor, J. & Pylyshyn Z. (2016). *Minds without Meanings*. MIT Press.

Forscher, P., Lai, C., R. Axt, J., Ebersole, C., Herman, M., Devine, P. & Nosek, B. (2016). "A Meta–Analysis of Change in Implicit Bias" https://psyarxiv.com/dv8tu/

Frankish, K. (2004). *Mind and Supermind*. CUP.

Frazier, B., Gelman, S., & Wellman, H. (2009). "Preschoolers' search for explanatory information within adult–child conversation", *Child Development*, 80, 1592–1611.

Freedman, M. (1994). "Atticus Finch—Right and Wrong", 45 *Ala. L. Rev*. 473.

Gaita, R. (2002). *The Philosopher's Dog*. Melbourne.

Galef, B. (1996). "Historical origins: the making of a science" In L. Houck and L. Drickamer (eds), *Foundations of Animal Behavior: Classic Papers with Commentaries*. Chicago University Press.

Gallagher, S. (2008). "Direct perception in the intersubjective context", *Consciousness and Cognition* 17 (2), 535–543.

Gallagher, S. & Hutto, D. (2008). “Understanding others through primary interaction and narrative practice” in Zlatev, Racine, Sinha & Itkonen (Eds.), *The Shared Mind* (17–38), John Benjamins Press.

Gallese, V. & Goldman, A. (1998). “Mirror neurons and the simulation theory of mind–reading”, *Trends in Cognitive Science*, 2(12) 493–501.

Gauker, C. (1988). "Objective Interpretationism", *Pacific Philosophical Quarterly* 69: 136–151.

Gauker, C. (1994). *Thinking Out Loud: An Essay on the Relation between Thought and Language,* PUP.

Gauker, C. (2003). *Words without Meaning,* MIT Press.

Geertz, C. (1973). *The Interpretation of Cultures*. Basic Books.

Gendler, T. (2008). "Alief and Belief", *Journal of Philosophy*, 105 (10):634–663.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2015). “How, whether, why: Causal judgments as counterfactual contrasts.” *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, Austin, TX, 782–787.

Gert, J. (2008). “What colors could not be: an argument for color primitivism” *Journal of Philosophy* (3):128–155

Glennan, S. (1996). “Mechanisms and the Nature of Causation”, *Erkenntnis*, 44:49–71.

Glüer, K. (2000). “Dreams and Nightmares”, in *Interpreting Davidson*, ed. P. Kotatko/P. Pagin/G. Segal, CSLI Publications: 53–74.

Godfrey–Smith, P. (1996). *Complexity and the Function of Mind in Nature*. CUP.

Godfrey–Smith, P. (2004). "On Folk Psychology and Mental Representation" *Representation in Mind: New Approaches to Mental Representation*, edited by H. Clapin, P. Staines, and P. Slezak, Elsevier.

Godfrey–Smith, P. (2005). "Folk Psychology as a Model" *Philosophers' Imprint* 5:6, 1–16.

Godfrey–Smith, P. (2016). *Other Minds*: *The Octopus, the Sea, and the Deep Origins of Consciousness*. Farrar, Straus, and Giroux.

Goldman, A. (1986). *Epistemology and Cognition*. HUP.

Goldman, A. (1989). "Interpretation Psychologized", *Mind & Language* 4:3, 161–185.

Goldman, A. (2006). *Simulating Minds*. OUP.

Gopnik, A. & Astington, J. (1988). "Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance–Reality Distinction" *Child Development,* Vol. 59, No. 1, 26–37.

Gopnik, A. & Wellman, H. (1992). "Why the Child's Theory of Mind Really is a Theory", *Mind & Language*, 7:1–2, 145–171.

Gopnik, A, Slaughter, V., & Meltzoff, A. (1994). "Changing your views: how understanding visual perception can lead to a new theory of mind", in Lewis & Mitchell (Eds.), *Children's Early Understanding of Mind: Origins and Development* (157–181), Taylor & Francis.

Gopnik, A. & Meltzoff, A. (1997). *Words, Thoughts and Theories*. MIT Press.

Gordon, R. (1986). "Folk Psychology as Simulation", *Mind & Language*, Vol 1:2, 158–171.

Gordon, R. (2002). “Simulation and Reason Explanation: The Radical View” in *Philosophical Topics*, Vol 29, 1&2, “The Philosophy of Alvin Goldman”.

Greenwald, A., McGhee, D., Schwartz, J. (1998). "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test", *Journal of Personality and Social Psychology*, 74 (6): 1464–1480.

Gregory, A. (2017). “Might Desires be Beliefs about Normative Reasons for Actions?” in Lauria & Deonna (eds.), *The Nature of Desire*, OUP.

Griffiths, P. (1997). *What emotions really are.* Chicago University Press.

Griffiths, P. (2004). “Emotions as natural and normative kinds” *Philosophy of Science*, 71, 901–911.

Gruber, T., Deschenaux, A., Frick, A., Clement, F. (2017). “Group Membership Influences More Social Identification than Social Learning or Overimitation in Children” *Child Development,* Early View.

Hacking, I. (1995). “The Looping Effects of Human Kinds” In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Symposia of the Fyssen Foundation. Causal cognition: A multidisciplinary debate* (pp. 351–394). OUP.

Hall, N. (2004). “Two Concepts of Causation” In John Collins, Ned Hall & Laurie Paul (eds.), *Causation and Counterfactuals* (225–276). MIT Press.

Hare, B., Call, J., Agnetta & Tomasello, M. (2000). “Chimpanzees know what conspecifics do and do not see. *Animal Behaviour* 59 (4): 771–785.

Hare, B., Call, J. & Tomasello, M. (2001). “Do chimpanzees know what conspecifics know? *Animal Behaviour* 61 (1): 139–151.

Hardin, C.L. (1988). *Color for Philosophers: Unweaving the Rainbow*. Hackett.

Harman, G. (1987). “(Nonsolipsistic) Conceptual Role Semantics,”, in Ernie LePore, ed., *New Directions in Semantics* (pp. 55–81). Academic Press.

Harman, G. (1999). “Moral philosophy meets social psychology: virtue ethics and the fundamental attribution error,” *Proceedings of the Aristotelian Society* 99, 315–331.

Harman, G. (2000). “The nonexistence of character traits,” *Proceedings of the Aristotelian Society* 100, 223–226.

Harris, P. (1992). “From simulation to folk psychology: the case for development”, *Mind & Language*, 7:1–2, 120–144.

Haslam, Nick (2006). “Dehumanization: an integrative review”, *Personality and Social Psychology Review* 10:3, 252–264

Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Critique*. OUP.

Hatfield, G. (1992). "Color Perception and Neural Encoding: Does Metameric Matching Entail a Loss of Information?", *PSA 1992: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, David Hull & Mickey Forbes (eds.), 1:492–504.

Hatfield, G. (2003). “Objectivity and Subjectivity Revisited: Color as a Psychobiological Property”. In *Colour Perception: Mind and the Physical World*, ed. by Rainer Mausfeld and Dieter Heyer (187–202). OUP.

Hatfield, G. (2009). *Perception & Cognition*. OUP.

Hatfield, G. (2014). "Cognition", in L. Shapiro (Ed.), *The Routledge Handbook of Embodied Cognition*

Hatfield, G. (2017). "Descartes as a Great Philosopher: Comprehensive Physics, Methodological Systematicity, and Mechanistic Embodiment" In S. Hetherington (Ed.), *What Makes a Philosopher Great?* (104–127), Routledge.

Haugleland, J. (1981). "Semantic Engines: An Introduction to Mind Design" in *Mind Design*. MIT Press.

Haugeland, J. (1982). "Weak supervenience", *American Philosophical Quarterly*, 19: 93–103.

Haugeland, J. (1998). *Having Thought: Essays in the Metaphysics of Mind*. HUP.

Haugeland, J. (2002). "Authentic Intentionality." In Matthias Scheutz, (Ed.), *Computationalism: New Directions* (pp. 159–174). MIT Press.

Heal, J. (1984). "Replication and functionalism" In Jeremy Butterfield (Ed.), *Language, Mind, and Logic* (135–150). CUP.

Heider, F. & Simmel, M. (1944). "An experimental study in apparent behavior". *The American Journal of Psychology*, 57, 243–259.

Heider, F. (1958). *The psychology of interpersonal relations*. Wiley.

Heil, J. (2004). *Philosophy of Mind: A Contemporary Introduction* Second Edition. Routledge.

Helton, G. (forthcoming). "Visually perceiving the intentions of others", *Philosophical Quarterly*.

Herschbach, M. (2015). "Direct social perception and dual process theories of mindreading", *Consciousness & Cognition* 36: 483–97.

Heyes, C. (2014). “Submentalizing: I Am Not Really Reading Your Mind” *Perspectives on Psychological Science* 9: 131–143.

Heyes, C. & Frith, C. (2014). “The Cultural Evolution of Mindreading”, *Science* 344, 1243091.

Heyes, C. (2015). “Animal mindreading: what’s the problem?” *Psychonomic Bulletin and Review* 22, 313–327.

Heyes, C. (2017). “Apes submentalize” *TICS,* 21, 1–2.

Heyes, C. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. HUP.

Hickling, A. & Wellman, H. (2001). “The emergence of children’s causal explanations and theories: Evidence from everyday conversation”, *Developmental Psychology*, 37, 668–683.

Hilbert, D. (1987). *Color and Color Perception: A Study in Anthropocentric Realism.* CSLI.

Hilbert, D. (1992). “What is color vision?”, *Philosophical Studies,* 68, 351–370.

Hochstein, E. (2016). “When does "Folk Psychology" Count as Folk Psychological?” *British Journal for the Philosophy of Science* 68:4, 1125–1147.

Hogrefe, G., Wimmer, H., & Perner, J. (1986). “Ignorance versus False Belief: a Developmental Lag in Attribution of Epistemic States”, 57:3, 567–582.

Horgan, T. & Tienson, J. (2002). “The Intentionality of Phenomenology and the Phenomenology of Intentionality” In David J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings* (520—533), OUP.

Hrdy, S. (2009). *Mothers and Others* Harvard University Press

Humphreys, P. (2016). *Emergence*. OUP.

Hurley, S. & Nudds, M. (2006). *Rational Animals?* OUP.

Humphrey, N. (1976). "The social function of the intellect" In *Growing Points in Ethology* ed. PPG Bateson and RA Hinde, 303–321. CUP.

Humphrey, N. (1978). "Nature's psychologists", *New Scientist*. 29:900–904.

Hunter, D. (2011). "Alienated belief", *Dialectica*, Vol 65, Iss 2, 221–240.

Hurvich, L. (1981). *Color Vision*. Sinauer Associates.

Hutto, D. (2008). *Folk Psychological Narratives.* MIT Press.

Hutto, D. (2009). "Lessons from Wittgenstein: Elucidating Folk Psychology", *New Ideas in Psychology*, 27, 197–212.

Hutto, D. (2011). "Presumptuous Naturalism: A Cautionary Tale" *American Philosophical Quarterly* Vol. 48, No. 2, Wittgenstein and Naturalism, 129–145.

Inagaki, K. & Hatano, G. (2002). *Young children's naïve thinking about the biological world*. Psychology Press.

Isen, A. & Levin, P. (1972). "Effect of Feeling Good on Helping: Cookies and Kindness", *Journal of Personality and Social Psychology* Vol. 21, No. 3, 384–388.

Jackson, F. (1977). *Perception: A Representative Theory*. Cambridge University Press.

Jackson, F. & Pettit, P. (1990). "In defence of folk psychology", *Philosophical Studies* 59: 31–54.

Jacobs, G. (1996). "Primate photopigments and primate color vision". *Proceedings of the National Academy of Science USA* 93:577–81.

James, W. (1896). "The Will to Believe", *The New World* 5: 327–347.

Jamieson, D. (2007). "What do Animals Think?" in Lurz, ed. *Philosophy of Animal Minds*, *CUP*, pp. 15–34.

Johnson, S.C., Booth, A., & O'Hearn, K. (2001). "Inferring the goals of a nonhuman agent". *Cognitive Development*, 16, 637–656.

Jones, E. & Harris, V. (1967). "The attribution of attitudes" *Journal of Experimental Social Psychology* 37: 398–412.

Kahan, D. (2015). "Climate–Science Communication and the Measurement Problem", *Political Psychology,* Vol 36, 1–43.

Kalish, C. (1998). "Reasons and causes: children's understanding of conformity to social rules and physical laws" *Child Development* 69:706–720.

Karg, K., Schmelz, M., Call, J., Tomasello, M. (2015). "Chimpanzees strategically manipulate what others can see." *Anim Cogn*. 18:1069–1076.

Karten, H. (2013). "Neocortical Evolution: Neuronal Circuits Arise Independently of Lamination", *Current Biology* 23 1:R12–15.

Kelemen, D. (1999). "The scope of teleological thinking in preschool children". *Cognition, 70,* 241–272.

Kelemen, D. (2003). "British and American children's preferences for teleological–functional explanations of the natural world". *Cognition, 8,* 201–221.

Kelemen, D., & DiYanni, C. (2005). "Intuitions about origins: Purpose and intelligent design in children's reasoning about nature". *Journal of Cognition and Development*, 6, 3–31.

Kelemen, D. & Rosset, E. (2009). "The Human Function Compunction: Teleological explanation in adults". *Cognition, 111*, 138–143.

Kelemen, D., Rottman, J. & Seston, R. (2013). "Professional physical scientists display tenacious teleological tendencies. Purpose–based reasoning as a cognitive default." *Journal of Experimental Psychology: General, 142*(4), 1074–1083.

Kelley, H. (1967). "Attribution Theory in Social Psychology" in D Levine, *Nebraska Symposium on Motivation* (15, 129–238), U Nebraska Press.

Kim, J. (2003). "The American Origins of Philosophical Naturalism", *Journal of Philosophical Research,* Vol 28, 83–98.

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). "The native language of social cognition". *Proceedings of the National Academy of Sciences, USA*, *104*, 12577–12580.

Kitcher, P. (1991). "Narrow Psychology and Wide Functionalism," in Richard Boyd, Phil Gasper, and J.D. Trout, (eds.), *The Philosophy of Science* (671–685). MIT Press.

Knausgaard, K.O. (2004). *A Time for Everything*. Archipelago Books.

Koffka, K. (1935). *Principles of Gestalt Psychology*. Harcourt, Brace and Company.

Köhler, W. (1947). *Gestalt Psychology: The Definitive Statement of the Gestalt Theory*. Liveright.

Kominsky, J., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). "Causal superseding". *Cognition*, 137, 196–209.

Korman, J. & Malle, B. (2016). "Grasping for Traits or Reasons? How People Grapple with Puzzling Social Behaviors", *Personality and Social Psychology Bulletin* Vol. 42(11) 1451–1465.

Kriegel, U. (2011). *The Sources of Intentionality.* OUP.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). "Great apes anticipate that other individuals will act according to false beliefs". *Science*, 354(6308), 110–114.

Krupenye, C., Kano, F., Hirata, S., Call, J., and Tomasello, M. (2017). "A test of the submentalizing hypothesis: Apes' performance in a false belief task inanimate control". *Communicative & Integrative Biology*. 10(4): e1343771.

Lauria, F. & Deonna, J. (2017) "Introduction: Reconsidering some Dogmas about Desire" in their *The Nature of Desire*. OUP.

Lazenby, R. (2015). *Michael Jordan: The Life*. Back Bay Books.

Lee, H. (1960). *To Kill a Mockingbird*. J. B. Lippincott & Co.

Lee, H. (2015). *Go Set a Watchman*. HarperCollins.

Leslie, A. (1987). "Pretense and Representation: The Origins of 'Theory of Mind'", *Psychological Review*, 94:4 412–426.

Leslie, S.J. & Lerner, A. (2016). "Generic Generalizations", *The Stanford Encyclopedia of Philosophy*.

Levin, J. (2004). "Functionalism", *The Stanford Encyclopedia of Philosophy*.

Lewis, D. (1966). "An Argument for the Identity Theory," *Journal of Philosophy*, 63: 17–25.

Lewis, D. (1972). "Psychophysical and theoretical identifications" *Australasian Journal of Philosophy*. 50(3), 249–258.

Lewis, D. (1973). *Counterfactuals*. Wiley–Blackwell.

Lewis, D. (1997). "Finkish Dispositions", *The Philosophical Quarterly* 47: 143–158.

Lillard, A. (1998). "Ethnopsychologies: Cultural Variations in Theories of Mind", *Psychological Bulletin*, 123, 3–32.

Livingston, J. (1997). "Metacognition", files.eric.ed.gov/fulltext/ED474273.pdf

Loar, B. (2003). "Phenomenal Intentionality as the Basis of Mental Content" In Hahn and Ramberg (eds.), *Reflections and Replies*. MIT Press.

Locke, J. (1690/1975). *Essay concerning human understanding*, ed. P.H. Nidditch. OUP.

Locke, D. & Pennington, D. (1982). "Reasons and other causes: their role in attribution processes" *Journal of Personality and Social Psychology* 42: 212–223.

Lomas, T. (2016). "Towards a positive cross–cultural lexicography", *The Journal of Positive Psychology*, 11:5, 546–558.

Lombrozo, T. (2010). "Causal–explanatory pluralism: How intentions, functions and mechanisms influence causal ascriptions" *Cognitive Psychology* 61 303–332.

Lombrozo, T., & Carey, S. (2006). Functional explanation and the function of explanation. Cognition, 99, 167–204.

Lombrozo, T. & Vasilyeva. N. (2017). "Causal explanation". In M. Waldmann (Ed.), *Oxford Handbook of Causal Reasoning.*

Lurz, R. (2011). *Mindreading Animals.* MIT Press.

Lycan, W. (1987). *Consciousness*. MIT Press.

Lycan, W. (1988). *Judgment and Justification*. MIT Press.

Lycan, W. (1996). *Consciousness and Experience*. MIT Press.

Macfarlane, J. (2014). *Assessment Sensitivity: Relative Truth and its Applications.* OUP.

Maibom, H. (2003). "The Mindreader and the Scientist", *Mind & Language* 18:3, 296–315.

Maibom, H. (2007). "Social Systems", *Philosophical Psychology*. 20: 557.

Maibom, H. (2009). "In Defence of (Model) Theory Theory", *Journal of Consciousness Studies*, 16, 360.

Malle, B. (1999). "How people explain behavior: a new theoretical framework" *Personality and Social Psychology Review* 3: 23–48.

Malle, B. (2004). *How the Mind Explains Behavior*. MIT Press.

Malle, B. (2011). "Time to Give up the Dogmas of Attribution: An Alternative Theory of Behavior Explanation" *Advances in Social Psychology*, 44, 298–338.

Malle, B. & Knobe, J. (1997). "The folk concept of intentionality" *Journal of Experimental Social Psychology* 33:101–121.

Malle, B. F., Knobe, J., O'Laughlin, M., Pearce, G., & Nelson, S. (2000). Conceptual structure and social functions of behavior explanations: Beyond person–situation attributions. *Journal of Personality and Social Psychology, 79,* 309–326.

Malle, B., Knobe, J., & Nelson, S. (2007). "Actor–observer asymmetries in behavior explanations: New answers to an old question" *Journal of Personality and Social Psychology, 93*, 491–514.

Mallon, R. (2006). "'Race': Normative, Not Metaphysical or Semantic." *Ethics* 116(3): 525–551.

Mameli, M. (2001). "Mindreading, mindshaping, and evolution" *Biology and Philosophy,* 16, 597–628.

Mandelbaum, E. (2013). "Against alief", *Philosophical Studies* vol. 165, no. 1, 197–211.

Mandelbaum, E. (2014). "Thinking is Believing", *Inquiry*, 2014 Vol. 57, No. 1, 55–96.

Mandelbaum, E. (2016). "Attitude, Inference, Association: On the Propositional Structure of Implicit Bias", *Nous* 50:3 (2016) 629–658.

Marcus, R.B. (1990). "Some revisionary proposals about belief and believing", *Philosophy and Phenomenological Research*, 50, 132–153.

Marcus, R.B (1995). "The anti–naturalism of some language centered accounts of belief", *Dialectica*, 49, 113–129.

Marr, D. (1982). *Vision*. MIT Press.

Marley–Payne, J. (2016). *Action First Attitudes*. Dissertation at MIT.

Marshall, J. (2018). "Is President Trump Mentally Ill? It Doesn't Matter", *Talking Points Memo,* Jan 6.

Matthews, R. (2007). *The Measure of Mind* OUP.

Matthews, R. (2011). “Measurement Theoretic Accounts of Propositional Attitudes“, *Philosophy Compass* 6/11: 828–841.

Mayer, A. & Träuble, B. (2013). “Synchrony in the onset of mental state understanding across cultures? A study among children in Samoa”. *Int. J. Behav. Dev*. 37, 21–28.

McDowell, J. (1986). “Singular Thought and the Extent of Inner Space” in (Pettit, P. and McDowell, J. eds.), *Subject, Thought and Context* (Clarendon Press, Oxford), 137–168.

McDowell, J. (1998). *Mind, Value and Reality*. HUP.

McGeer, V. (2007). “The regulative dimension of folk psychology”, *Folk Psychology Re Assessed*, D. Hutto, M. M. Ratcliffe, (Eds.), Springer, 137–156.

McNaughton, D. (1988). *Moral Vision*. Blackwell.

Mele, A. (2003). *Motivation and Agency*. OUP.

Mele, A. (2017). *Aspects of Agency*. OUP.

Menzel, R., Steinmann, E., de Souza, J., Backhaus, W. (1988). "Spectral Sensitivity of Photoreceptors and Colour Vision in the Solitary Bee, Osmia Rufa". *Journal of Experimental Biology* 136 (1): 35–52.

Menzies, P. (2015). “The Role of Counterfactual Dependence in Causal Judgments”, in Christoph Hoerly, Teresa McCormack & Sarah Beck (eds.), *Understanding Counterfactuals, Understanding Causation*. OUP.

Miller, C. (2013) *Moral Character: An Empirical Theory*. OUP.

Millikan, R. (1984). *Language, Thought and Other Biological Categories*, MIT Press.

Millikan, R. (1989). “Biosemantics”, *Journal of Philosophy*, 86: 281–97.

Millikan, R. (1993). *White Queen Psychology and Other Essays for Alice*, MIT Press

Millikan, R. (2000a). *On Clear and Confused Ideas: An Essay about Substance Concepts* Cambridge University Press.

Millikan, R. (2000b). "Reading Mother Nature's Mind," in *Dennett’s Philosophy: a Comprehensive Assessment,* D, Ross, A. Brook and D. Thompson, eds. (Cambridge MA: MIT Press/Bradford), 55–75.

Mitchell, S. (2005). “Anthropomorphism and Cross–Species Modelling”, in Daston & Mittman (eds.) *Thinking with Animals: New Perspectives on Anthropomorphism* (Columbia), 100–108.

Mölder, B. (2010). *Mind Ascribed: An Elaboration and Defense of Interpretivism*. John Benjamins Publishing.

Mollon, J. (1989). ““Tho' she kneel'd in that place where they grew...” The uses and origins of primate colour vision” *Journal of Experimental Biology* 146:21–38.

Monroe, A. & Malle, B. (2017). “Two paths to blame: Intentionality directs moral information processing along two distinct tracks.” *Journal of Experimental Psychology: General*, Vol 146(1),123–133.

Morgan, C.L. (1883a). “Suicide of Scorpions”, *Nature* 27, 313–314

Morgan, C.L. (1883b). “Scorpion Suicide”, *Nature* 27, 530

Morgan, C.L. (1884). “Instinct”, *Nature* 29, 370–374

Morgan, C.L. (1887). “Scorpion Virus”, *Nature* 35, 535

Morgan, C.L. (1891). *Animal Life and Intelligence.* Ginn & Company.

Morgan, C.L. (1894). *An Introduction to Comparative Psychology*. Walter Scott Publishing Company.

Morgan, C.L. (1903). *An Introduction to Comparative Psychology.* Second Edition. Walter Scott Publishing Company.

Morgan, C.L. (1930a). "Autobiography of C. Lloyd Morgan", in Murchison, Carl. (Ed.), *History of Psychology in Autobiography* Vol. 2, Springer, 237–264.

Morgan, C.L. (1930b). *The Animal Mind*. Edward Arnold.

Morton, A. (2003). *The importance of being understood: folk psychology as ethics.* Routledge.

Morton, A. (2007). "Folk Psychology Does Not Exist". In Hutto, D. & Ratcliffe, M. (eds.), *Folk Psychology Re–Assessed*. Kluwer/Springer Press 211–221.

Nagel, J. (2011). "Knowledge as a mental state", *Oxford Studies in Epistemology* Volume 4, Chapter 10.

Neander, K. (2017). *A Mark of the Mental: In Defense of Informational Teleosemantics*. MIT Press.

Nichols, S. & Stitch, S. (2003). *Mindreading*. OUP.

Nisbett, R. E., Krantz, D. H., Jepson, D., & Kunda, Z. (1983). "The use of statistical heuristics in everyday reasoning", *Psychological Review*, *90*, 339–363.

Nolfi, K. (2015). "How to Be a Normativist About the Nature of Belief," *Pacific Philosophical Quarterly* 96 (2):181–204.

O'Brien, K., V. Slaughter, C. C. Peterson (2011). "Sibling influences on theory of mind development for children with ASD". *J. Child Psychol. Psychiatry* 52, 713–719.

Oswald, F. & Mitchell, G. & Blanton, H. & Jaccard, J. & Tetlock, P. (2013). "Predicting Ethnic and Racial Discrimination: A Meta–Analysis of IAT Criterion Studies. Journal of personality and social psychology". 105.

Papineau, D. (1984). "Representation and Explanation", *Philosophy of Science*, 51: 550–72.

Papineau, D. (1987). *Reality and Representation*. Blackwell.

Papineau, D. (1993). *Philosophical Naturalism*. Blackwell.

Peacocke, C. (1992). *A Study of Concepts*. MIT Press.

Penn, D. & Povinelli, D. (2007). "On the lack of evidence that non–human animals possess anything remotely resembling a "theory of mind"" *Philos Trans R Soc Lond B Biol Sci.;* 362(1480): 731–744.

Perez–Zapata, D., Slaughter, V., and Henry, J.D. (2016). "Cultural Effects on Mindreading" *Cognition* 146, 410–414.

Perner, J. (1990). *Understanding the Representational Mind*. MIT Press.

Perner, J. (2010). "Who took the cog out of cognitive science? Mentalism in an era of anti–cognitivism" in P. A. Frensch, & R. Schwarzer (Eds), *Cognition and Neuropsychology: International Perspectives on Psychological Science*, Volume 1 (pp. 241–61). Psychology Press.

Perner, J. & Roessler, J. (2009). "Teleology and Causal Understanding in Children's Theory of Mind" in J. Aguilar, & A. Buckareff (Eds), *Causing Human Action: New perspectives on the causal theory of action* (pp. 199– 228) MIT Press.

Piccinini, G. (2006). "Computation without Representation", *Philosophical Studies*, 137:2.

Place, U.T. (1956). "Is consciousness a brain process?", *British Journal of Psychology* 47:44–50.

Place, U.T. (1996). Four chapters of *Dispositions: A Debate* (with D.M. Armstrong and C.B. Martin), ed. Tim Crane, Routledge.

Place, U.T. (2004). *Identifying the Mind*, ed. G. Graham. OUP.

Polger, T. (2004). *Natural Minds*. MIT Press.

Polger, T. & Shapiro, L. (2017). *The Multiple Realization Book.* OUP.

Povinelli, D. & Vonk, J. (2003). "Chimpanzee Minds: Suspiciously Human?" *TICS* 7:4, 157–160.

Premack, D. & Woodruff, G. (1978). "Does the chimpanzee have a theory of mind?" *Behavioral and Brain Sciences* 4 (4):515–629.

Price, H. H. (1969). *Belief.* London: Allen & Unwin.

Proust, J. (2015). *The Philosophy of Metacognition*. OUP.

Pryal, K.R.G. (2010). "Walking in Another's Skin: Failure of Empathy in to Kill a Mockingbird", *Harper Lee's To Kill a Mockingbird: New Essays* (M. Meyer, ed.) Scarecrow Press, 174–189.

Putnam, H. (1967). “The mental life of some machines” in Hector–Neri Castaneda (ed.), *Intentionality, Minds, and Perception.* Wayne State University Press.

Putnam, H. (1975). "The Meaning of meaning". *Minnesota Studies in the Philosophy of Science* 7:131–193.

Pyers, Gollan, and Emmorey (2009). “Bimodal bilinguals reveal the source of tip–of–the–tongue–states” *Cognition*, 112, 323–329.

Pyle, A. (1999). *Great Philosophers in Conversation*. Routledge.

Pylyshyn, Z. (1980). "Cognition and Computation". *Behavioral and Brain Sciences* 3, 111–32.

Pylyshyn, Z. (1984). *Computation and Cognition*. MIT Press.

Quilty–Dunn, J. & Mandelbaum, E. (2017). “Against dispositionalism: belief and cognitive science”, *Philosophical Studies*. Online First.

Quine, W.V.O. (1960). *Word and Object*. MIT Press.

Quine, W.V.O. (1969). *Ontological Relativity and Other Essays*. Columbia University Press.

Quine, W.V.O. (1981). *Theories and things*. Belknap Press.

Quine, W.V.O. (1990). “Three indeterminacies” in Barret and Gibson (ed.) *Perspectives on Quine*, Wiley–Blackwell, 1–16.

Quine, W.V.O. (1992). *The Pursuit of Truth*. HUP.

Rabin, J., Houser, B., Talbert, C., & Patel, R. (2016). “Blue–Black or White–Gold? Early Stage Processing and the Color of ‘The Dress’”, *PLoS ONE* 11(8).

Ramsey, W., Stich, S., & Garon, J. (1990). “Connectionism, eliminativism and the future of folk psychology” In William Ramsey, Stephen P. Stich & D. Rumelhart (eds.) *Philosophy and Connectionist Theory*. Lawrence Erlbaum. 499–533.

Ratcliffe, M. (2006). *Rethinking Commonsense Psychology*. Palgrave Macmillan.

Regan, B.C, C. Julliot, B. Simmen, F. Viénot, P. Charles–Dominique, Mollon, J.D. (2001). "Fruits, foliage and the evolution of primate colour vision" P*hil. Trans. R. Soc. Lond.* B 2001 356 229–283.

Rejesus, R., Mutuc–Hensley, M., Mitchell, P., Coble, K., and Knight, T. (2013). “U.S. Agricultural Producer Perceptions of Climate Change*” Journal of Agricultural and Applied Economics*, 45, 4:701–718.

Reutlinger, A. (2012). “Getting rid of interventions”, *Studies in the History and Philosophy of Science Part C: 43 (4):* 787–795.

Reutlinger, A. (2017). “Explanation beyond causation? New directions in the philosophy of scientific explanation”, *Philosophy Compass*. Online First.

Riddle, C. (2013). “Defining Disability: Metaphysical Not Political” in *Medicine, Health Care, & Philosophy* 16, no. 3: 377–384.

Riggs, K. & Peterson, D. (2000). “Counterfactual thinking in pre–school children: mental state and causal inferences” in *Children’s Reasoning and the Mind*, Psychology Press, 87–99.

Roessler, J. & Perner, J. (2013). “Teleology: Belief as Perspective” in Baron–Cohen, Tager–Flusberg, & Lombardo (Eds.), *Understanding Other Minds* (35–50), OUP.

Romanes, G. (1888). *Mental Evolution in Man*. Kegan Paul, Trench & Co.

Rosenberg, A. (2016). "The Genealogy of Content or the Future of an Illusion", *Philosophia* 43(3): 537–547.

Ross, L. (1977). "The intuitive psychologist and his shortcomings: distortions in the attribution process", *Advances in Experimental Social Psychology*, 10, 174–221. Academic Press.

Rosset, E. (2008). It's no accident: Our bias for intentional explanations. *Cognition*, 108, 771–780.

Rowbottom, D. (2016). "How Might Degrees of Belief Shift? On Actions Conflicting With Professed Beliefs", *Philosophical Psychology* 29(5), 732–742.

Ruffman, T., Slade, L., & Crowe, E. (2002). "The relation between children's and mothers' mental state language and theory–of–mind understanding. *Child Development, 73*, 734–751.

Russell, B. (1910). "Knowledge by acquaintance and knowledge by description", *Proceedings of the Aristotelian Society* 11: 108–28.

Russell, B. (1921). *The Analysis of Mind*. George Allen & Unwin.

Russell, J. (2017). "Kyrie Irving believes the Earth is flat. It is not." *The Washington Post* https://www.washingtonpost.com/news/early–lead/wp/2017/02/17/kyrie–irving–believes–the–earth–is–flat–it–is–not.

Ryle, G. (1949). *The Concept of Mind*. Chicago University Press.

Ryle, G. (1953). *Dilemmas*. Cambridge University Press.

Ryle, G. (1966). "Jane Austen and the Moralists" in *Collected Papers, Vol 1.* Hutchinson.

Ryle, G. (1971a). *Collected Papers*, Volume 1. London: Hutchinson.

Ryle, G. (1971b). *Collected Papers*, Volume 2. London: Hutchinson.

Ryle, G. (1979). *On Thinking*, K. Kolenda (ed.), Oxford: Blackwell.

Saidel, E. (2009). "Animal Thoughts" in Robert Lurz (ed.) *Animal Minds*. CUP.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. PUP.

Scott, R. & Baillargeon, R. (2017). "Early False Belief Understanding", *Trends in Cognitive Science*. Online First.

Schaafsma, S., Pfaff, D., Spunt, R., Adolphs, R., (2015). "Deconstructing and reconstructing theory of mind", *Trends in Cognitive Sciences*, 19(2) 65–72.

Schaffer, J. (2007). "Knowing the Answer" *Philosophy and Phenomenological Research*, Vol. LXXV No. 2.

Schlaffke, L., Golisch, A., Haag, L., Lenz, M., Heba, S., Lissek, S., Schmidt–Wilcke, T., Eysel, U., Tegenthoff, M. (2015). "The brain's dress code: How The Dress allows to decode the neuronal pathway of an optical illusion", *Cortex*, 73: 271–275.

Schroeder, T. (2003). "Donald Davidson's Theory of Mind is Non–Normative", *Philosophers' Imprint*, Vol 3, No. 1, 1–14.

Schult, C & Wellman, H. (1997). "Explaining human movements and actions: children's understanding of the limits of psychological explanation" *Cognition, 62*, 291–324.

Schwitzgebel, E. (2001). "In–between believing" *Philosophical Quarterly*, 51, 76–82.

Schwitzgebel, E. (2002). "A Phenomenal, Dispositional Account of Belief" *Nous*, 36, 249–275.

Schwitzgebel, E. (2010). "Acting Contrary to Our Professed Beliefs, or the Gulf Between Occurrent Judgment and Dispositional Belief", *Pacific Philosophical Quarterly*, 91, 531– 553.

Schwitzgebel, E. (2013). "A Dispositional Approach to Attitudes: Thinking Outside of the Belief Box", in Nikolaj Nottelmann (ed.), *Belief* (Palgrave).

Schwitzgebel, E. (2018). "It's not just one thing, to believe there's a gas station on the corner", *The Splintered Mind*. Retrieved from http://schwitzsplinters.blogspot.com /2018/02/its–not–just–one–thing–to–believe.html.

Schwitzgebel, E. (in draft). "The Pragmatic Metaphysics of Belief", retrieved from http://www.faculty.ucr.edu/~eschwitz/SchwitzAbs/PragBel.htm.

Schwitzgebel and Meyers–Schulz (2013). "Knowing That P Without Believing That P", *Nous*, 47, 371–384.

Searle, J. (1992). *The Rediscovery of the Mind*. MIT Press.

Searle, J. (1994). "Animal Minds", *Midwest Studies in Philosophy,* XIX, pp. 206–219.

Searle, J. (1995). *The Construction of Social Reality*. The Free Press.

Sehon, S. (2005). *Teleological Realism*. MIT Press.

Sellars, W. (1956). "Empiricism and the Philosophy of Mind", *Minnesota Studies in the Philosophy of Science, Volume I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Herbert Feigl and Michael Scriven, eds., 253–329.

Sellars, W. (1962). "Philosophy and the Scientific Image of Man," in *Frontiers of Science and Philosophy*. (Robert Colodny, ed.). University of Pittsburgh Press.

Shapiro, L. (2008). "How to Test for Multiple Realization," *Philosophy of Science* 75: 514–525.

Shapiro, L. (2018). "Reduction redux," *Studies in History and Philosophy of Science Part A,* Volume 68, 10–19.

Shaw–Williams, K. (2013) "The Social Trackways Theory of the Evolution of Human Cognition", *Biological Theory.* Online First.

Shelbie, T. (2016). *Dark Ghettoes: Injustice, Dissent, and Reform*. HUP.

Shettleworth, S. (2007). *Cognition, Evolution and Behavior.* OUP.

Shettleworth, S. (2010). "Clever animals and killjoy explanations in comparative psychology" *Trends in Cognitive Sciences* 14:11, 477–481.

Simons, D. & Keil, F. (1995). "An abstract to concrete shift in the development of biological thought: the *insides* story", *Cognition* 56, 129–163.

Skinner, B.F. (1938). *The Behavior of Organisms*. Appleton & Company.

Slaughter, V. & Perez–Zapata, D. (2015). "Cultural Variations in the Development of Mind Reading", Volume 8, Issue 4, 237–241.

Smart, J.C.C. (1959). "Sensations and Brain Processes", *Philosophical Review* 68: 141–156.

Smart, J.C.C. (1961). "Colours", *Philosophy* 36:128–42.

Smart, J.C.C. (2000). "The Mind/Brain Identity Theory", *The Stanford Encyclopedia of Philosophy.*

Smith, J.D. (2009). “The study of animal metacognition” *Trends in Cognitive Science* 13:9, 389–396.

Smith, J. (2013). “The Phenomenology of Face–to–Face Mindreading”, *Philosophy and Phenomenological Research* 90:2, 274–293.

Smith, Z. (2016). “On Optimism and Despair”, *The New York Review of Books,* Dec 22.

Sober, E. (1985). "Panglossian Functionalism and the Philosophy of Mind" in P. Machamer (ed.) Naturalistic Epistemology. *Synthese,* 64, 165–94.

Sober, E. (1998). "Morgan's Canon." In C. Allen and D. Cummins (eds.), *The Evolution of Mind*, OUP, 224–242.

Sober, E. (2005). “Comparative Psychology Meets Evolutionary Biology: Morgan’s Canon and Cladistic Parsimony”, in Daston & Mittman (eds.) *Thinking with Animals: New Perspectives on Anthropomorphism* (Columbia), 85–99.

Sommers, F. (2009). “Dissonant beliefs”, *Analysis*, 69:2, 267–274.

Southgate, V., Senju, A., & Csibra, G. (2007). “Action anticipation through attribution of false belief by 2–year–olds”, *Psychological Science*, 18:7, 587–592.

Spaulding, S. (2015a). “On Direct Social Perception”, *Consciousness and Cognition* 36:472–482.

Spaulding, S. (2015b). “Mind Misreading”, *Philosophical Issues*, Vol 26:1, 422–440.

Spaulding, S. (forthcoming). “Do You See What I See? How Social Differences Influence Mindreading”. *Synthese*.

Spencer, Q. (2014). "A Radical Solution to the Race Problem", *Philosophy of Science* 81:5, 1025–1038.

Sterelny, K. (1990a). "Animals and Individualism", *Vancouver Studies in Cognitive Science* 1, 323–339.

Sterelny, K. (1990b). *The Representational Theory of Mind*. Blackwell.

Sterelny, K. (2012). *The Evolved Apprentice.* MIT Press.

Stich, S. (1979). "Do animals have beliefs?" *Australasian Journal of Philosophy* 57:15–28.

Stich, S. (1981). "Dennett on Intentional Systems", *Philosophical Topics*, 12:1, 39–62.

Stich, S. (1983). *From Folk Psychology to Cognitive Science.* MIT Press.

Stich, S. (1990). "Review of *The Intentional Stance*", *Ethics*, 100, 4, 891–892.

Stich, S. (1997). *Deconstructing the Mind*. OUP.

Strawson, G. (1994). *Mental Reality*. MIT Press.

Tajfel, H. (1974). "Social Identity and intergroup behavior", *Soc. Sci, Info.,* 13:2, 65–93.

Tanney, J. (2013). *Rules, Reason, and Self–Knowledge.* HUP.

Tarski, A. (1944). "The Semantic Conception of Truth and the Foundations of Semantics", *Philosophy and Phenomenological Research*, 4: 341–376.

Taumoepeau, M. & Ruffman, T. (2006). "Mother and infant talk about mental states relates to desire language and emotion understanding". *Child Development* 77, 465–481.

Taumoepeau, M. & T. Ruffman. (2008). "Stepping stones to others' minds". *Child Development* 79, 284–302.

Taylor, P.C. (2000). "Appiah's Uncompleted Argument: W.E.B. Du Bois and the Reality of Race", *Social Theory and Practice* 26(1): 103–128.

Terrace, H. & Son, L. (2009). "Comparative Metacognition" *Current Opinions in Neurobiology* 19(1):67–74.

Thomas, R. (1998). "Lloyd Morgan's Canon". In G. Greenberg and M. Haraway (eds), *Comparative Psychology: A handbook.* Garland.

Thomas, R. (2001). "Lloyd Morgan's Canon: A History of Misrepresentation" *History & Theory of Psychology Eprint Archive*.

Thompson, E. (1995). *Color Vision.* Routledge.

Thomson, E. & Piccinini, G. (2018). "Neural representations observed", *Minds and Machines* 28 (1): 191–235.

Titelbaum, M. (forthcoming). *Fundamentals of Bayesian Epistemology*. OUP.

Tobia, K., Newman, G., & Knobe, J. (2017). "Water is and is not H20", https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2933100.

Tolman, E.C. (1932). *Purposive Behavior in Animals and Men*. Random House.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition.* HUP.

Tumulty, M. (2014). "Managing Mismatch Between Belief and Behavior," *Pacific Philosophical Quarterly* 95 (3) 261–92.

van Fraasen, B.C. (1980). *The Scientific Image*. OUP.

Verheggen, C. (2007). "Triangulating with Davidson", *Philosophical Quarterly* 57 (226):96–103.

Verheggen, C. (2016). "From Triangulation to Semantic Normativity", Ch. 2 of *Donald Davidson's Triangulation Argument* (co–authored with Robert Myers). Routledge.

Wallisch, P. (2017). "Illumination assumptions account for individual differences in the perceptual interpretation of a profoundly ambiguous stimulus in the color domain: 'The Dress'", *Journal of Vision* 17:5 1–14.

Watson, J., Gelman, S. & Wellman, H. (1998). "Young children's understanding of the non–physical nature of thoughts and the physical nature of the brain" *British Journal of Developmental Psychology, 16*, 321–335.

Webster, M. (2015). "Individual Differences in Color Vision", *Handbook of Color Psychology*, A.J. Elliot, M.D. Fairchild, and A. Franklin (Eds.), Cambridge University Press, 197–215.

Weisberg, M. (2007). "Three Kinds of Idealization" *The Journal of Philosophy* 104 (12):639–659.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. OUP.

Wellman, H. (2014). *Making Minds: How Theory of Mind Develops*. OUP.

Wellman, H., Cross, D., & Watson, J. (2001). "A meta–analysis of theory–of–mind development: the truth about false belief", *Child Development*, 72, 655–684.

Wellman, H. & Gelman, S. (1998). "Knowledge acquisition in foundational domains", in (D. Kuhn & R. Siegler eds.), *Handbook of child psychology Vol 2., Cognition, perception, and language*, Wiley, 523–573.

Westra, E. (2016). "Spontaneous Mindreading: A Problem for the Two–Systems Account" *Synthese*. Online first. 1–23.

Westra, E. (2018a). "Character and theory of mind: an integrative approach", *Philosophical Studies*. Online First.

Westra, E. (2018b). "Stereotypes, theory of mind, and the action–prediction hierarchy". *Synthese*. Online First.

Wilkerson, W. (1997). "Real Patterns and Real Problems," *Southern Journal of Philosophy* 97 (4):557–70.

Willison, R. (2017). *The Ethics of Meaning*. Dissertation at the University of Pennsylvania.

Wilson, G. (1989). *The Intentionality of Human Action*. Stanford University Press.

Wimmer, H. & Perner, J. (1983). "Beliefs about beliefs", *Cognition* 13, 103–128.

Woodward, J. (2003). *Making Things Happen*. OUP.

Wu, S. & Keysar, B. (2014). "The Effect of Culture on Perspective Taking", *Psychological Science* vol. 18, no. 7, 600–606.

Wynne, C. (2007). "What are Animals? Why Anthropomorphism is Still not a Scientific Approach to Behavior" *Comparative Cognition and Reviews*, 2, 125–135.

Yablo, S. (1995). "Singling out properties" *Philosophical Perspectives*, 9, 477–502.

Yalowitz, S. (2012). "Anomalous Monism", *The Stanford Encyclopedia of Philosophy.*

Zahavi, D. (2011). "Empathy and Direct Social Perception: A Phenomenological Proposal", *Review of Philosophy and Psychology* 2(3): 541–558.

Zawidzki, T. (2013a). *Mindshaping*. MIT Press.

Zawidzki, T. (2013b). "Commentary on Butterfill and Apperly's 'How to Construct a Minimal Theory of Mind?'" *Brains Blog* http://philosophyofbrains.com/wp–content/uploads/2013/11/zawidzki.pdf

Zawidzki, T. (2015). "Dennett's Strategy for Naturalizing Intentionality: An Innovative Play at Third Base", *Philosophia* 43:3, 593–609.