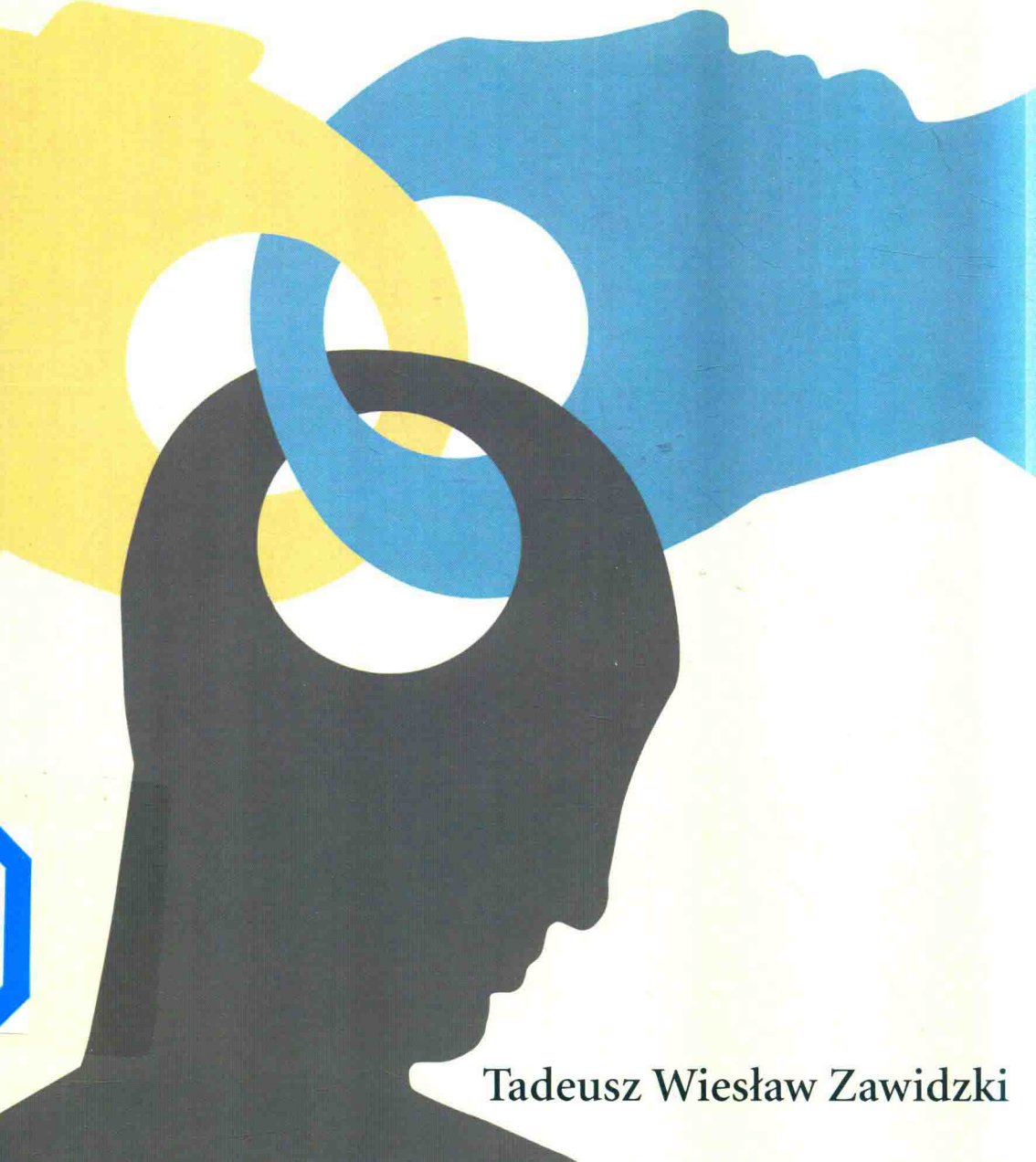
Mindshaping

A New Framework for Understanding Human Social Cognition



Tadeusz Wiesław Zawidzki

Mindshaping

A New Framework for Understanding Human Social Cognition

Tadeusz Wiesław Zawidzki

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email special_sales@mitpress.mit.edu or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Stone Sans and Stone Serif by the MIT Press. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Zawidzki, Tadeusz.
Mindshaping : a new framework for understanding human social cognition / Tadeusz Wiesław Zawidzki.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-262-01901-9 (hbk. : alk. paper)
1. Social perception. 2. Social psychology. 3. Social cognitive theory. I. Title.
HM1041.Z33 2013
302'.12—dc23
2012038062

10 9 8 7 6 5 4 3 2 1

Mindshaping

For Zofia Zyta Zawidzka, Tadeusz Witold Zawidzki, and Elżbieta Aniela Zawidzki: first-rate mindshapers all.

Who can guess what convoluted nonsense may be brewing in the minds of our fellow men, sometimes along with deep intuitions.

—Czesław Miłosz, *A Year of the Hunter* (1994)

The uniformity that unites us in communication and belief is a uniformity of resultant patterns overlying a chaotic subjective diversity of connections between words and experience. Uniformity comes where it matters socially. ... Different persons growing up in the same language are like different bushes trimmed and trained to take the shape of identical elephants. The anatomical details of twigs and branches will fill the elephantine form differently from bush to bush, but the overall outward results are alike.

—Willard Van Orman Quine, *Word and Object* (1960)

The reality of a man's life isn't found only where he is. It's also found in the other lives that shape his—first of all, the lives of those he loves ... but also the lives of unknown others—powerful or downtrodden—fellow citizens, policemen, professors, invisible companions in mines and factories, diplomats and dictators, religious reformers, artists who create myths that govern our behavior—all told, humble representatives of the sovereign accidents that reign over even the most orderly existence.

—Albert Camus, *Discours de Suède* (1958)

Preface

The central thesis of this book is that mindshaping (Mameli, 2001) is the linchpin of the human sociocognitive syndrome. Much of what follows is devoted to elucidating the notion of mindshaping and discussing its many instantiations in human social experience. The basic idea is that we are distinguished in our social behavior from other species primarily by our sophisticated, complex, diverse, and flexible capacities to shape each other's minds. Examples include sophisticated imitation, pedagogy, irresistible conformism, norm institution, cognition and enforcement, and narrative self- and group constitution. The claim that mindshaping constitutes the linchpin of the human sociocognitive syndrome amounts to the following: without mindshaping, none of the other components of distinctively human social cognition—sophisticated language, sophisticated and pervasive cooperation, and even sophisticated mindreading—would be possible. This distinctively human sociocognitive syndrome would not have evolved without mindshaping, and were mindshaping somehow removed from it now, the other components would disappear as well.

This thesis seeks to displace the reigning metaphor in the sciences of human social cognition: mindreading. According to this metaphor, our distinctive sociocognitive profile can be traced entirely to our mindreading virtuosity: what sets us apart is that we are natural psychologists of unparalleled skill, capable of representing each other's mental states with a reliability, accuracy, and sophistication unmatched by any other species. Mindreading means different things to different people, and a major burden of what follows, especially chapter 1 and chapter 3, is making clear distinctions among the varieties of mindreading and identifying those that are distinctive of human beings. We share many mindreading capacities with nonhuman animals. For this reason, these mindreading capacities cannot be what, in the received view, sets us apart. But we have good evidence that one variety of mindreading—which has been the focus of most

philosophical and psychological research into human social cognition—is distinctive of human beings: the accurate attribution of propositional attitudes like belief and desire. Many researchers simply assume that the distinctively human sociocognitive syndrome depends most fundamentally on this capacity. This claim is the specific target of my critical agenda: I argue that the attribution of full-blown propositional attitudes cannot have evolved before sophisticated practices of mindshaping aimed at making us easily interpretable to each other. It is likely that sophisticated mindshaping coevolved in the human lineage with improved versions of sociocognitive capacities that we *share* with nonhuman primates, such as tracking the goals of conspecifics and anticipating the rationally and informationally constrained behavioral means they select to achieve them. However, as I argue, even highly sophisticated versions of such behavior tracking do not amount to the attribution of full-blown propositional attitudes. The capacity to attribute such mental states depends on, and had to await, the evolution of sophisticated mindshaping practices, especially linguistic practices like norm institution and narrative self- and group constitution.

The distinction between the mindreading- and mindshaping-focused understandings of human social cognition cannot be captured in terms of simple empirical tests. No crucial experiment can vindicate one understanding at the expense of the other. In fact, the mindshaping-focused understanding embraces many of the same empirical results as the mindreading-focused understanding, although it interprets them differently and attaches greater importance to some and lesser importance to others. If scientific frameworks were judged only by their ability to explain or predict specific empirical results, then this would jeopardize the interest of the project I pursue in this book. However, as philosophers of science have long appreciated (Kuhn, 1977), there is more to scientific virtue than just empirical adequacy. Theories and paradigms are also evaluated on the basis of nonempirical virtues like simplicity, coherence with the rest of what science tells us about the world, and suggestiveness of new directions for research. Different metaphors afford different interpretations of the same data, and some interpretations are simpler, more consistent with other disciplines, and more suggestive than others. In what follows, I make the case that mindshaping affords interpretations of empirical data that are superior to those afforded by mindreading along these nonempirical dimensions.

Mindreading evokes the following picture of human social cognition. Our epistemic relations to each other are no different from our epistemic relations to the nonhuman world. Other people are mysterious phenomena

animated by unobservable causes that are completely independent of our attempts to understand them. Individual human beings must learn to infer these causes to better predict and control other human beings, just as is the case with nonhuman phenomena. They do so by attempting to represent these unobservable causes, and they succeed to the extent that they represent them accurately. Our social accomplishments—pervasive, institutionalized cooperation and coordination, language, and so on—can all be explained in terms of such individualized sociocognitive accomplishments. In short, according to the mindreading metaphor, distinctively human social cognition is conceptualized as an individual accomplishment, involving the accurate representation of independently constituted, unobservable mental causes of behavior.

In contrast, mindshaping, as I formulate it here, treats our epistemic relations to each other as very different from those we hold to the nonhuman world. Our frameworks for interpreting human behavior succeed so well because they are simultaneously frameworks for *regulating* human behavior (McGeer, 2007). The states we attribute to each other in interpretation are not independent of our interpretive frameworks because those frameworks shape the states our minds are likely to enter. We do not first discover independent facts about what causes behavior, and then inform our attempts to control each other with such discoveries. Our social accomplishments are not by-products of individualized cognitive feats, like the discovery of some unobservable causal factor. Rather, through a form of "group selection,"[1] simultaneously interpretive and regulative frameworks that support our social accomplishments, including pervasive, institutionalized cooperation and coordination, language, and so on, have evolved. In the mindshaping metaphor, distinctively human social cognition is conceptualized as a group accomplishment, involving simultaneously interpretive and regulative frameworks that function to shape minds, which these frameworks can then be used to easily and usefully interpret. In what follows, I hope to illustrate, through a detailed reading of a variety of empirical work, that this conceptualization brings greater simplicity and unity to a greater variety of empirical results from the sciences of social cognition than does the mindreading conceptualization. I also hope to give the reader a taste of how suggestive of new research the mindshaping metaphor can be.

Mindshaping inevitably evokes a discredited image of the human mind: that it is entirely the product of shaping by environmental factors—regimes of punishment and reward instituted by society. This behaviorist understanding of the human mind has rightfully been discredited since

the cognitivist revolution of the 1960s. The evidence that human minds bring to their learning environments highly complex cognitive mechanisms is overwhelming. However, this in no way contradicts the thesis that mindshaping constitutes the linchpin of the human sociocognitive syndrome. The various kinds of mindshaping all require sophisticated cognitive mechanisms in both "shapees" and shapers. This does not, however, imply that these cognitive mechanisms involve sophisticated mindreading, especially the attribution of full-blown propositional attitudes. What follows departs from common contemporary presuppositions not in the assumption that human social learning presupposes complex cognitive equipment but in the assumption that this cognitive equipment must include a capacity to attribute full-blown propositional attitudes. On the contrary, I argue that the capacity to accurately attribute full-blown propositional attitudes is parasitic, in phylogeny and today, on prior capacities to shape minds.

A related worry: some might balk at the term "mindshaping." To many, "mind" connotes the enduring, underlying mechanisms that implement all intelligent human behavior, including social learning and other forms of mindshaping. Such "core cognition" (Carey, 2009) is not subject to shaping or learning; rather, it makes them possible. Relative to core cognition, the kinds of dispositions that *are* subject to modification—for example, public language lexicons and accents, musical and other aesthetic preferences, skills like tool use or game play, and so on—seem too ephemeral to count as significant parts of the mind. Perhaps "behavior shaping" is a better term.[2] However, our commonsense notion of "mind" surely connotes more than enduring core cognition. For example, no one thinks that quotidian mind*reading* succeeds in discovering and tracking the activity of core cognition; it took decades of careful scientific research to identify such fundamental components of the mind. Similarly, it is no part of my claim, nor need it be, that quotidian mind*shaping* aims to alter core cognition. The varieties of mindshaping that I explore here all aim to alter behavioral dispositions. It is hard to see how altering behavioral dispositions is possible without altering minds in some sense. Hence "mindshaping" is an appropriate term.

A final caveat: the view I defend is entirely neutral on questions of mechanism. I have no strong views on whether or not the competencies I describe are implemented by innate, special-purpose symbolic modules, or by general-purpose, experience-driven neural networks, or by some hybrid of the two, or by some other alternative. Compelling evidence supports all these approaches to understanding the "how?" of human cognition, and

it is likely that the brain employs a diversity of mechanisms with varied architectures. I focus here on the "why?" of human social cognition: the functions that sustain our sociocognitive capacities.

My strategy is largely empirical. I think a close reading of contemporary research into human social cognition across a variety of disciplines makes a strong case for the thesis that mindshaping is the key innovation behind the human sociocognitive syndrome. However, this is more than just a survey. I deploy philosophical analysis to make clear the distinctions between varieties of both mindreading and mindshaping and to make explicit precisely what sorts of cognitive capacities each different variety presupposes. For example, many theorists slide easily from the claim that distinctively human social phenomena, like culture, language, and pervasive cooperation, require a form of social cognition more sophisticated than mere behavioral generalization to the claim that they require the capacity to attribute full-blown propositional attitudes like beliefs and desires. But this claim ignores a variety of alternatives that are more sophisticated than mere behavioral generalization yet do not qualify as propositional attitude attribution. Many theorists also make facile assumptions about what beliefs and desires are, and consequently underestimate the cognitive sophistication presupposed by their accurate attribution. This has important consequences for assumptions about the viability of propositional attitude attribution as our primary means of navigating the social world. The definition of mindshaping also calls for some philosophical sophistication, as it is not apparent how one can intelligently shape a mind without first representing it accurately, which would make sophisticated mindshaping parasitic on sophisticated mindreading. So the reader should expect a philosophically responsible close reading of a variety of empirical literature pertaining to human social cognition, marshaled in support of the mindshaping-as-linchpin hypothesis. I turn now to a brief overview of what I intend to accomplish in each of the seven chapters.

Chapter 1 provides an overview of the entire project and, in particular, a detailed unpacking of what I call "the human sociocognitive syndrome."[3] I argue that human social cognition is distinguished from other varieties by four broadly related capacities: sophisticated mindreading, sophisticated mindshaping, extremely diverse and pervasive cooperation, and structurally complex and semantically flexible language. In the received view, sophisticated mindreading forms the linchpin of this sociocognitive syndrome: without our capacity to accurately attribute propositional attitudes, none of the other components are possible. Chapter 1 sets the agenda for

the rest of the book, arguing that contrary to the received view, the human sociocognitive syndrome relies on sophisticated mindshaping rather than on sophisticated mindreading.

The goal of chapter 2 is to clarify what mindshaping is. To this end, I defend a philosophical analysis of the notion, grounded in Ruth Millikan's (1984) notion of biological proper function. Basically, any mechanism the proper function of which is getting a target mind to match a model in certain respects counts as a mindshaping mechanism. I then deploy this understanding to taxonomize the varieties of mindshaping observed in both human and nonhuman populations. It turns out that human mindshaping is distinguished from nonhuman varieties in four ways: (1) the sheer variety of the respects in which it can make targets match models; (2) the fact that only humans find matching models intrinsically motivating, rather than merely a means to further ends; (3) the socially distributed nature of much human mindshaping, for example, the involvement of teachers; and (4) the ontological status of models used in some human mindshaping, for example, fictional characters or idealized agents. The main burdens of chapter 2 are to (1) provide a clear definition of mindshaping, (2) survey the empirical literature to illustrate its varieties, and (3) show how human mindshaping can be distinguished from nonhuman varieties without assuming that it presupposes sophisticated mindreading, like propositional attitude attribution.

Chapter 3 motivates the project of the book by means of a critical discussion of the received view of human social cognition, that is, the view that sophisticated mindreading, and especially propositional attitude attribution, forms the basis of most distinctively human sociocognitive accomplishments. I raise a variety of objections to this assumption, all based on the observation that, due to the *holism* of the propositional attitudes, the relations between observable behavior and propositional attitudes are too complex for propositional attitude attribution to help with the accurate and timely prediction of behavior. This explains why propositional attitude attribution does not occur among nonhuman species, even highly social and intelligent ones like contemporary great apes. However, it raises a deep puzzle about the evolution of human mindreading: how and why did our ancestors develop this capacity, given its relative intractability as a tool for behavioral prediction? I end chapter 3 by suggesting that only with prior, sophisticated mindshaping, ensuring cognitive homogeneity in populations of likely interactants, can distinctively human mindreading be reliable and timely enough to evolve.

Chapter 4 defends a detailed phylogenetic account tracing the evolution of distinctively human mindshaping in human prehistory. I begin with Sterelny's (2003, 2007, 2012) proposal that cooperation explains our distinctive evolutionary trajectory. I then argue that, for various reasons, sophisticated mindreading and, in particular, propositional attitude attribution cannot help explain the evolution of cooperation. On the contrary, drawing on a variety of empirical evidence, as well as models proposed by philosophers and evolutionary game theorists, I argue that various mechanisms of mindshaping aimed at homogenizing human populations best explain the roots of human cooperation. In particular, mindshaping dispositions and practices, like irresistible conformism, pedagogy, and norm institution and enforcement, make possible a form of group selection without which distinctively human cooperation would not be possible. Furthermore, such mindshaping dispositions and practices can also explain human virtuosity at forming plural subjects, which is crucial to explaining how we succeed in coordinating on cooperative projects in which we are already motivated to engage.

Chapter 5 addresses one of the central objections to the mindshaping-as-linchpin hypothesis. Many of the distinctively human mindshaping mechanisms and practices taxonomized in chapter 2, and employed to explain human phylogeny in chapter 4, presuppose competence in a structurally complex and semantically flexible public language. However, it is commonly assumed that human language use presupposes sophisticated mindreading and, in particular, the capacity to attribute full-blown propositional attitudes. One reason for this is that utterances of public language typically suggest multiple interpretations, and understanding the communicative intentions of one's interlocutors seems to be the only way of eliminating such indeterminacy (Sperber & Wilson, 1995, 2002). Another reason is that influential theories of the phylogeny of human language presuppose the capacity to attribute propositional attitudes (Bickerton, 1998, 2000, 2002; Origgi & Sperber, 2000). I address the first reason by distinguishing between full-blown propositional attitude attribution and what Dennett (1987) calls adopting the "intentional stance." While full-blown propositional attitude attribution requires mindreading, I argue that (some of Dennett's own characterizations of it notwithstanding) adopting the intentional stance does not.[4] I then argue that adopting the intentional stance, without any kind of mindreading, can better mitigate interpretive indeterminacy in conversation than full-blown propositional attitude attribution. I address the second reason by defending an alternative model of the phylogeny of human language.

Selection for increasingly complex ritualistic signaling systems, aimed at advertising reliability and competence in coordination on cooperative projects, can explain how contemporary human language became structurally complex, without assuming a prior capacity to attribute propositional attitudes.

Chapter 6 extends the defense of the project by addressing influential interpretations of recent evidence concerning the ontogeny of human mindreading. In these interpretations, human children acquire the capacity to attribute propositional attitudes before learning a language (Scott & Baillargeon, 2009; Buttelmann et al., 2009), and they cannot learn a language without first being able to attribute communication-relevant propositional attitudes to adult models (Bloom, 2000, 2002). Extending the strategy of chapter 5, I argue that the data are better interpreted as showing that very young human children adopt versions of the intentional stance without attributing full-blown propositional attitudes. When this capacity is supplemented with infant dispositions to treat stereotyped adult communicative behavior as pedagogical (Csibra & Gergely, 2009), the acquisition of language can be explained without assuming that prelinguistic infants can attribute full-blown propositional attitudes. I end chapter 6 with a brief survey of persuasive evidence that human infants and children are default "mindshapers" and "mindshapees" from a very young age.

An implication of chapters 2 through 6 is that there seems to be very little for full-blown propositional attitude attribution to do. Adopting the intentional stance, when supplemented with pervasive, diverse, and sophisticated mindshaping practices, should be enough to explain most distinctively human sociocognitive feats. How and why, then, did our species develop the capacity to attribute full-blown propositional attitudes? In particular, what use would sensitivity to a distinction between mental reality and behavioral appearance have served if the sophisticated tracking of behavioral patterns made possible by the intentional stance, supplemented by mindshaping, was sufficient for highly reliable behavioral prediction? Chapter 7 argues that full-blown propositional attitude attribution, and the distinction between behavioral appearance and mental reality that it presupposes, first evolved to serve a *social* rather than an *epistemic* function. That is, rather than enabling more accurate representations of others' thoughts, propositional attitude attribution allowed individuals to maintain, rehabilitate, or diminish each other's social status in the wake of counternormative behavior, like reneging on explicit discursive commitments. As Jerome Bruner (1990) argues, narratives appealing to an agent's propositional attitudes often seem to perform an exculpatory or justificatory

function, normalizing apparently deviant behavior by identifying reasons for it. Recent work in social psychology (Malle et al., 2007) confirms this idea. In chapter 7, I draw on such evidence to defend the view that propositional attitudes function primarily as components of justificatory and regulative, self-constituting narratives that ensure that likely interactants' expectations are appropriately and mutually adjusted. I conclude the book with a brief survey of the wide swath of psychological and philosophical inquiry that the notion of mindshaping has the potential to integrate.

Acknowledgments

This book has been a long time coming. Along the way, I've profited greatly from the generous help of numerous friends and colleagues. The final product is immeasurably better due to their (often unwitting) input and support. Of course, I'm entirely to blame for any remaining flaws.

I'm most grateful to J. Robert Thompson and Shannon Spaulding for their detailed and copious feedback on earlier drafts of this project. Not only did they patiently read very rough chapter drafts, but their insightful comments and suggestions regarding research of which I was unaware helped move the project in more promising directions. And as if this weren't enough, they helped organize an online reading group of the first complete draft, enlisting an incredibly energetic, generous, and talented group of young philosophers and cognitive scientists whose regular feedback through the spring and summer of 2010 dramatically improved the quality of the final product while helping me sustain the energy I needed to complete it. I am grateful to Olle Blomberg, Ryan DeChant, Mitch Herschbach, Richard Moore, Josh Shepherd, and Michael Wilby for their enthusiastic participation in the reading group and their many extremely helpful suggestions. I feel fortunate to have had this many creative and insightful thinkers devote so much of their energy to helping me with this project.

Another online forum helped germinate the ideas in this book: Pete Mandik generously agreed to host discussion of a draft of the paper from which this project sprang on his *Brain Hammer* blog in the fall of 2006. The project started off on the right foot thanks to comments from participants in the blog discussion.

I must also express my deep gratitude to the Department of Philosophy at George Washington University for the supportive and stimulating environment they provided during the drafting of this book. Most of my colleagues have read all or parts of earlier drafts, providing extremely helpful comments at departmental brown bags, as well as in less formal contexts. I

thank David DeGrazia not just for detailed comments on a number of chapters but for being a true mentor when I needed it most. I'm also deeply grateful to Eric Saidel, Jeff Brand, Michele Friend, Gail Weiss, Paul Churchill, and Bill Griffiths for reading and responding to early drafts of bits of the book and challenging me in ways that made it better. I am exceedingly lucky to work in such a supportive and exceptionally collegial department.

I am also fortunate to be at an institution that values and supports interdisciplinary collaboration. This has made possible regular contact with fellow members of the Mind-Brain-Evolution Cluster and the Mind-Brain Institute at George Washington University. I want especially to thank Chet Sherwood and Francys Subiaul for their friendship, support, and constant updates on the latest in comparative psychology and neuroscience.

Others who have provided helpful feedback to earlier versions of the ideas presented here include Robert Lurz, Javier Gomez-Lavin, Amy Kind, Cynthia MacDonald, Whit Schonbein, Marc Slors, and Şerife Tekin. I also thank Rik Hine for drawing my attention to the very apt quotation by Czesław Miłosz employed as one of the epigraphs.

I would also like to acknowledge the support of Dan Dennett. Besides the tremendous intellectual debt that this project owes his groundbreaking work in philosophy of mind, I am grateful for some highly useful feedback Dan found time to give me regarding my interpretation of some of his views, though I'm not sure I take them in a direction that he fully endorses. Dan is truly a giant on whose shoulders I try, perilously, to perch. He has been an invaluable source of support, not just in this project but also throughout my career.

I also owe a huge intellectual debt to Dan Hutto, whose *Folk Psychological Narratives* (2008) crystallized my own growing reservations about orthodox approaches to human social cognition. Besides his important intellectual influence, Dan has been supportive in more practical ways as well and was instrumental in making this project a reality.

I would also like to thank four anonymous reviewers employed by the MIT Press to evaluate the penultimate draft. Their comments were rigorous, detailed, and challenging. I hope I have gone some way toward satisfactorily addressing their worries in the final version. Either way, their efforts undoubtedly improved the final product. The whole MIT Press team deserves kudos for their extreme professionalism and patient support. I'm especially grateful to my former editor, Tom Stone, for believing in me early in the project, to my current editor, Philip Laughlin, for his continuing support and professionalism, to the manuscript editor, Judy Feldmann, for her very prompt and professional management of the production process, and