Mindshaping: A New Framework for Understanding Human Social Cognition

Tadeusz Wiesław Zawidzki

The MIT Press

From The MIT Press

MITCogNet

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

1 The Human Sociocognitive Syndrome

1 Our Sociocognitive "Linchpin": Mindreading or Mindshaping?

Despite vigorous debate about a variety of issues, a consensus is growing among cognitive scientists interested in the evolutionary origins of human social cognition. The consensus concerns how to characterize the human sociocognitive syndrome, or the set of sociocognitive capacities that distinguishes our species from our closest primate relatives. This consensus is about *what* has evolved in the human lineage; it is the background to dramatic disagreement about two *how*-questions: how these capacities evolved, and how they are implemented in the brain. Consensus has it that human social cognition is distinguished from nonhuman social cognition by four broadly related capacities: sophisticated mindreading, sophisticated mindshaping,[1] pervasive cooperation, and structurally complex and flexible symbolic communication.

Cognitive scientists have also reached a strong though somewhat less pervasive consensus on the relations of phylogenetic dependency among these four capacities. According to many researchers, sophisticated mindreading is the linchpin holding the human sociocognitive syndrome together. Without sophisticated mindreading, it is claimed, sophisticated mindshaping, pervasive cooperation, and structurally complex and flexible symbolic communication would not be possible (Humphrey, 1980; Tooby & Cosmides, 1995, xvii; Baron-Cohen, 1999; Leslie, 2000, p. 61; Mithen, 2000; Sperber, 2000; Dunbar, 2000, 2003, 2009; Siegal, 2008, p. 22).[2] This largely unquestioned background assumption is often asserted without argument to motivate research into mental state attribution. For example, in a recent paper claiming to show that even seven-month-old infants have some understanding of others' mental states, Kovács et al. (2010) begin as follows:

> Humans are guided by internal states such as goals and beliefs. Without an ability to infer others' mental states, society would be hardly imaginable. Social interactions,

> from collective hunting to playing soccer to criminal justice, critically depend on the ability to infer others' intentions and beliefs. Such abilities are also at the foundation of major evolutionary conundra. For example, the human aptitude at inferring mental states might be one of the crucial preconditions for the evolution of the cooperative social structure in human societies. (1830)

According to this view, the other components of the human sociocognitive syndrome depend on sophisticated mindreading, especially propositional attitude attribution. We are able to shape each other's minds so much more effectively than nonhuman primates, through imitation, pedagogy, and norm enforcement, for example, because our capacity for mindreading is so much more sophisticated than theirs. Just as shaping the physical world, that is, engineering, improves dramatically when guided by more sophisticated and accurate theories of physical domains, so human mindshaping is a dramatic improvement over nonhuman mindshaping because it is guided by a more sophisticated and accurate theory of mind. We succeed in cooperating in a dramatically wider variety of endeavors, of a dramatically greater complexity, involving far greater numbers of interactants than nonhuman primates, because we are far better mindreaders, and hence far better at anticipating each other's behavior. Finally, our capacity for structurally complex and flexible symbolic communication both is made possible by sophisticated mindreading and makes possible even more sophisticated mindreading. Language is the paradigm of human communication, though there are other uniquely and universally human forms, like music and dance used in ritual. And according to the received view, language is made possible by distinctively human mindreading, especially the recognition of communicative intentions, and also makes possible a kind of mindreading unmatched by any other species: we can learn each other's thoughts, to seemingly arbitrary degrees of precision, by communicating them to each other.

The principal goal of this book is to articulate and defend an alternative picture of the relations of phylogenetic dependency between the four capacities that comprise the human sociocognitive syndrome. In this alternative, sophisticated mindreading is not the linchpin that holds the whole syndrome together. Sophisticated mindreading is unlikely to evolve or remain stable in social contexts that have not already been structured by sophisticated forms of mindshaping. On the contrary, sophisticated mindreading is possible only in social contexts comprising minds that have already been shaped to be easily, mutually interpretable. In the alternative I defend, sophisticated mindshaping, presupposing mindreading capacities little more sophisticated than those available to our closest nonhuman cousins,[3] emerged in the hominid lineage leading to modern humans because such mindshaping made possible dramatic improvements in cooperation. Structurally complex and flexible symbolic communication evolved, first and foremost, as a mindshaping, cooperation-enhancing tool. This then made sophisticated mindreading possible, including the attribution of full-blown propositional attitudes, like beliefs and desires with linguistically specifiable contents.[4]

This characterization of the distinction between the received and alternative views is somewhat oversimplified. For example, everyone agrees that we use structurally complex, flexible symbolic communication to shape our social environment in ways that are conducive to cooperation. Proponents of the received view acknowledge that the four components of the human sociocognitive syndrome likely *coevolved*. Improvements in mindreading made possible improvements in mindshaping, cooperation, and communication. These then likely set up social structures that selected for improved mindreading, setting off a coevolutionary spiral (Pinker, 2003). However, in the received view, sophisticated mindreading is the first and most important step in this process. Evolution of mindshaping, cooperation, and communication in more humanlike directions is impossible without prior evolution of more humanlike mindreading.

The received view follows directly from the following assumptions: (1) humanlike mindreading is both possible and biologically advantageous in populations not yet characterized by humanlike mindshaping, cooperation, and communication, and (2) these latter three capacities are impossible without humanlike mindreading. To defend the mindshaping-first alternative, I argue that (1) more humanlike mindreading is *not* possible or biologically advantageous in populations lacking humanlike mindshaping, cooperation, and communication, and (2) these capacities can exist before humanlike mindreading. Making these arguments is the principal task of this book. In the course of making them, I also illustrate the variety, sophistication, pervasiveness, and biological uniqueness of human mindshaping and argue that many everyday interpretive practices widely assumed to serve a primarily mindreading function actually serve a primarily mindshaping function. In the rest of this chapter, I make the various concepts and distinctions central to my arguments clearer by briefly reviewing the latest literature on distinctively human mindreading, mindshaping, cooperation, and communication. First I turn to a brief note on the methodological challenges faced by any account of the evolution of and relations of phylogenetic dependency between the components of the human sociocognitive syndrome.

2 The Nature and Value of Evolutionary Hypotheses about Cognition

The most obvious objection to the project I have outlined concerns the value of engaging in speculations as unconstrained by data as hypotheses about the evolution of cognitive traits. Given that cognitive traits do not fossilize, it seems impossible to avoid the pitfalls of "just-so" storytelling. Despite this, we have seen an explosion of recent speculation concerning the evolution of the four components of the human sociocognitive syndrome. Since this book seeks to contribute to this literature, I will not spend much time defending an endeavor the value of which it takes for granted. However, I will outline a two-part defense against criticism of speculation about the evolutionary origins of cognitive capacities. First, I will review the surprisingly rich variety of indirect evidence that can keep such speculation responsible. Second, I will outline some of the important roles that evolutionary speculation can play in suggesting hypotheses and constraining research in other areas of cognitive science.

Before I turn to these arguments, let me clarify the approach to evolutionary explanation that I assume in this book. The core notion is that of an "evolutionarily stable strategy" (ESS) (Maynard-Smith, 1982; Lessard, 2006). According to this perspective, phylogenetic hypotheses aim to explain why some trait came to invade a population of interbreeding organisms and how it has remained stable against invasion by mutants that lack the trait.[5] Here is how I intend to apply this perspective to the human sociocognitive syndrome. The question of phylogenetic priority and determination between mindreading and mindshaping becomes the following. We begin with the last common ancestor we share with contemporary chimpanzees (henceforth LCA). We then attempt to infer, using all available evidence, what kinds of mindreading and mindshaping were likely possible for that species. Next we take into account other known, relevant features of their ecology and ask the following question: is it more likely that more humanlike mindreading invaded such a population before more humanlike mindshaping, or is the opposite more likely? In addition, we ask whether the presence of one or the other in a population would make invasion by the other more likely. Given this framework, one can phrase the central thesis of this book with more precision. I argue that it is much more likely that humanlike mindshaping invaded populations of the LCA or its descendants than that humanlike mindreading did. I also argue that once such mindshaping was stable, it made more humanlike mindreading more likely to invade. The details of the argument turn on the unviability of humanlike mindreading in populations without mindshaping practices that make individuals more easily interpretable to each other. It also appeals to ecological factors that made pervasive and sophisticated cooperation central to the success of our precursors, and the claim that such cooperation is impossible without the kind of humanlike mindshaping that makes humanlike mindreading viable.

Although we have no direct evidence concerning the cognitive capacities of the LCA and other human precursors, we do have a variety of indirect evidence. Most obviously, we can consult the fossil record. Although neural tissue and cognitive capacities do not fossilize, the skeletal remains that do fossilize often provide decent evidence of cognitive capacity. Most obviously, the size and shape of fossilized hominid skulls give some indication of cognitive capacity. Although brain size is not a perfect index of cognitive capacity, absolute brain size is the best anatomical correlate of humanlike cognition among primates (Sherwood et al., 2008, pp. 444–446). Skull endocasts can also reveal gross patterns of brain organization (p. 442). Skull size and shape can be used to reliably date dramatic expansions in brain size in the hominid lineage. These can be correlated with other evidence to suggest hypotheses about what drove such dramatic expansions. For example, if there is no evidence of dramatic ecological change correlating with such expansions, this is reason to think that they were driven by social factors, for example, expanding group sizes or competition with other groups (Sterelny, 2007). Furthermore, such fossil data can also be correlated with archaeological data, for example, early stone tools and their geographical distribution, to suggest hypotheses about the correlation between neural expansion and extent and quality of tool use, as well as social complexity sufficient to allow for long-range migration.

Another interesting source of evidence concerns interactions between cranial and postcranial changes. For example, there is strong evidence of bipedality in fossilized remains of extinct hominid species going back as far as 4.4 million years (Lovejoy et al., 2009). But radical cranial expansion imposes severe costs on bipedal mammals because bipedality constrains the size of the birth canal through which large-headed neonates must pass (DeSilva & Lesnik, 2008). Such costs must be added to the already considerable metabolic costs of supporting a large brain (Aiello & Wheeler, 1995). We can infer from such facts that the benefits of having a larger brain at a certain point in hominid history must have been considerable indeed. The evidence suggests that brain expansion among our precursors began to accelerate significantly with the rise of *Homo erectus* about 1.8 million years ago (Holloway et al., 2004). This suggests that dramatic, perhaps runaway selection pressures for increased cognitive capacities must have arisen in this time period. Whatever these selection pressures were, they must have overpowered the considerable costs of birthing large-headed neonates and nutritionally meeting their extreme metabolic needs. According to one plausible hypothesis, the birthing problem was partially mitigated because the neonates of large-headed, bipedal hominids were born with incompletely developed brains, in small, partially collapsed skulls, as human neonates are (Aiello, 1996). This suggestion has significant implications for hypotheses about our precursors' socioecology and related cognitive capacities.

The premature-brain-at-birth hypothesis suggests that, as adult brain size expanded, infants were born increasingly helpless. Provisioning such infants with requisite calories would require alloparenting—mothers would have to be helped by other members of a population, including fathers and blood relatives. This would require a degree of cooperation unmatched among extant nonhuman hominids. In addition, prolonged childhood would increase the opportunities for, and importance of, mindshaping practices like guided imitation and pedagogy. That increasing degrees of neural growth would occur postnatally while infants were exposed to contingent environmental stimuli—both physical and social—suggests that hominids in our lineage developed minds that were far more plastic and sensitive to local contingencies than those of other primates. While species whose brains are almost fully wired at birth have neurocognitive profiles almost entirely determined by genes, and therefore by transgenerationally stable aspects of the environment, species like our own, in which much brain wiring occurs postnatally, would have far more locally sensitive neurocognitive profiles. This has potentially profound implications for mindreading. The task of tracking the mental properties of humans and human precursors is bound to be exceptionally difficult if the course of neural development is driven to such a degree by contingent environmental regularities. Such populations are likely to display far less cognitive and behavioral homogeneity than populations in which most neural development is complete prenatally and hence fixed genetically. This inevitably makes mindreading more challenging in human beings and their immediate precursors than in other primates. I will expand on these hypotheses in the rest of the book, especially chapter 4, which examines the evolution of distinctively human mindshaping. The foregoing is merely a taste of how far even indirect fossil evidence can take us in developing responsible hypotheses about the phylogenesis of the human sociocognitive syndrome.

The second most important source of evidence constraining phylogenetic hypotheses is the archaeological record. This is admittedly a very noisy signal. For example, McBrearty and Brooks (2000) argue that Eurocentric prioritization of archaeological exploration, and demographic biasing that results from denser populations leaving more artifacts, have led to dramatic misinterpretations of the archaeological record. According to these misinterpretations, distinctively human cognition is a relatively recent phenomenon, emerging suddenly about fifty thousand years ago. McBrearty and Brooks argue that recent excavations in Africa contradict this hypothesis, suggesting a gradual emergence of humanlike cognition and social organization over the last 200,000 to 300,000 years. Despite such controversies, as archaeological evidence accumulates, it supports responsible conjecture about the phylogenesis of the human sociocognitive syndrome. For example, artifacts found in locations at great distances from the materials of which they are made provide good evidence of elaborate, long-distance trade routes, requiring complex social networks (McBrearty & Brooks, 2000). Ornamentation of artifacts and evidence of bodily ornamentation, like red ocher, indicate the marking of ethnic groups, an important signal of group selection for cooperative traits (Sterelny, 2003, 2012, pp. 54–55). Finally, the geographic diversity of technology and ornament is an important signal of the human sociocognitive syndrome. Despite their wide geographic range, protohuman hominids, like *Homo erectus*, had an extremely uniform, stereotyped tool kit, suggesting important social and cognitive differences from our own species (Mithen, 1996).

Another major source of evidence constraining phylogenetic hypotheses about human cognitive capacities comes from comparative psychology, neuroanatomy, and genetics. Humans differ in significant ways from their genetically closest living nonhuman relatives. Genetic evidence can support remarkably precise dating of the emergence of cognitive traits. For example, neuropsychological studies have shown the FOXP2 gene to play an important role in human language use (Lai et al., 2001). Important differences between humans and other extant primates at this gene locus have been identified and dated to the last 200,000 years (Enard et al., 2002). Significant behavioral differences also distinguish human beings from their closest nonhuman relatives. Comparative psychology has shown that all nonhuman primates lack the human capacities for belief attribution, fine-grained and flexible imitation, sophisticated cooperation, and complex language (Tomasello et al., 2005). This makes it likely that the LCA lacked these capacities, too, and gives us a good idea of the basic sociocognitive profile from which ours evolved. Finally, there are also significant neural differences between human beings and their closest nonhuman relatives. The most obvious and well-supported difference is in absolute brain size

(Sherwood et al., 2008). But many have also argued that there are significant differences in the sizes of different brain regions relative to the rest of the brain and to overall body mass. For example, the visual cortex appears to take up a far smaller proportion of the human brain than the brains of nonhuman primates, while the opposite is true of the prefrontal cortex (p. 441). Also, lateralization appears to be more pronounced in the human brain than in the brains of other primates (pp. 432–433).

The final major source of evidence constraining phylogenetic hypotheses about the human sociocognitive syndrome comes from comparative studies of different human populations. Although the universality of a trait is not an infallible marker of its genetic basis, it certainly makes it more likely that the trait is the product of genes selected for in human evolution. There is little doubt that the four components of the human sociocognitive syndrome are present in all nonpathological human populations.

Thus, although responsible speculation about the phylogenesis of the human sociocognitive syndrome is extremely challenging, requiring the integration of information across numerous disciplines, it is not impossible. At least four distinct kinds of evidence provide substantial constraints. Furthermore, formulating responsible hypotheses about the evolutionary roots of human cognition is not only possible; it plays an important role in cognitive science. The foundational metaphor of contemporary cognitive science likens the mind/brain to an information processor or computer. And to explain the behavior of an information processor, one must first specify the task in which it is engaged. This is the point of the three-level methodology that Marr (1982) proposes for computational cognitive science. The top, or computational, level specifies what task a cognitive process is trying to accomplish. The middle, or algorithmic, level specifies, at an abstract level, how the cognitive process accomplishes the task, that is, what kinds of algorithms it uses, and over what kinds of representations these algorithms are defined. The lowest, or implementational, level specifies, in neural terms, how the brain implements these algorithms.

When it comes to biological information processors, like the human brain, there appear to be only two ways of addressing questions at the top, computational level. Like Marr (1982), one can appeal to intuition or introspection to specify in what task some cognitive process is engaged. For example, Marr argues that human vision consists in mapping a two-dimensional retinal image onto a cognitive representation of a three-dimensional scene. This certainly seems like the task that vision accomplishes. However, there are many reasons to think that our introspectively grounded intuitions about the tasks in which our brains are engaged can be systematically misleading (Clark, 1997, epilogue). Indeed, Churchland et al. (1994) criticize Marr's model of vision on precisely these grounds. They argue that a better guide to the tasks in which human and other biological cognitive processes are engaged is evolutionary pedigree.[6] The human visual system evolved primarily to guide locomotion, not to construct detailed representations of three-dimensional scenes. According to Churchland et al., the human visual system is a messy, unprincipled "bag of tricks" for guiding an organism around ecologically plausible terrains. The growing evidence that human vision is implemented in at least two independently functioning pathways—the dorsal pathway devoted to locating objects and the ventral pathway devoted to classifying them (Milner & Goodale, 2006)—supports this view, as well as general skepticism about introspection-guided intuition as a method for determining the tasks in which brains are engaged. There is no intuitive or introspective reason to expect that the brain's capacity to locate objects should be completely dissociable from its capacity to classify them. But from an evolutionary perspective, this makes perfect sense. Even very simple organisms need a quick, automatic capacity to visually locate objects, since this is necessary for effective locomotion through a cluttered environment. More sophisticated capacities for classifying objects according to various properties, like color, are likely more recent evolutionary innovations—new tricks added to the already existing collection.

Thus if one wants to avoid the pitfalls of introspection- and intuition-guided speculation about what tasks the human mind/brain performs, the only alternative appears to be responsible speculation about what it evolved to do. This is one important role that evolutionary hypotheses can play in cognitive science. I think many assumptions about human social cognition have been based on intuition-based hypotheses about what social cognition is for, similar to Marr's hypothesis about what vision is for. It seems intuitively compelling that one cannot accomplish any social tasks—communicating, cooperating, mindshaping—without first constructing accurate models of the minds of one's interactants. Thus the principal task of human social cognition must be accurate mindreading. The analogy to Marr's claim about vision is instructive. It is also intuitive that vision cannot successfully guide navigation through, or manipulation of, the physical environment without first constructing an accurate, detailed model of its three-dimensional structure. However, as Churchland et al. argue, naturally evolved cognitive systems seldom respect our intuitions. There may be far messier and less principled, though far more efficient and robust, solutions discovered by naturally evolved systems. Such possibilities can be uncovered only with the help of responsible speculation about phylogenesis. I argue that when this perspective is applied to the human sociocognitive syndrome, sophisticated mindshaping turns out to be a far more important component of human sociocognitive competence than sophisticated mindreading.

Besides enabling cognitive scientists to frame the question of what cognitive processes are for without relying entirely on introspectively guided intuition, phylogenetic hypotheses also suggest explanations for otherwise puzzling experimental results. For example, it is well established that human beings are much better at reasoning about social norm violation than about nonsocial contingencies, even when these are formally similar (Cosmides, 1989). When asked whether there is a rule that a card with a vowel on one side must have an even number on the other, subjects systematically fail to consider cards they know to have odd numbers on one side, even though turning such cards over could show that the rule does not hold if even one has a vowel on the other side. In contrast, subjects have no trouble engaging in such reasoning when the topic is norm violation. For example, if the rule states that if one is drinking alcohol, one must be twenty-one or older, subjects automatically respond by checking both whether all those drinking alcohol are twenty-one *and* whether all those younger than twenty-one are not drinking alcohol. This is a paradigm-setting result for evolutionary psychology: it is taken to indicate that the human mind includes a cognitive module dedicated to detecting cheaters, something that is highly plausible given widely accepted hypotheses about prevalent social circumstances in human evolution (Cosmides & Tooby, 1992). Counterintuitive empirical results concerning similarities and differences in sexual preference between the genders (Carruthers, 2006, pp. 41–42), amount of care devoted to offspring depending on various contextual factors (pp. 42–43), and even improved memory for arbitrary words in survival-related versus survival-neutral contexts (Nairne & Pandeirada, 2008) are also easily explained from an evolutionary perspective. There is an increasing body of empirical evidence concerning human psychology that makes sense only relative to responsible hypotheses about the evolutionary raisons d'être of human cognitive capacities.

3 Mindreading

"Mindreading" (Nichols & Stich, 2003) has become a term of art in the literature on human social cognition. It refers to a phenomenon also called "mentalizing" (Goldman, 2006) or exercising one's "theory of mind" (Premack & Woodruff, 1978). These terms are used rather loosely, referring to a diverse assortment of phenomena. In its most inclusive sense, "mindreading" refers to the exercise of a cognitive capacity aimed at anticipating the behavior of other agents, on the basis of *some* appreciation of the mental states responsible for it. But this characterization obscures significant differences among the phenomena that different theorists call "mindreading." Since the goal of this book is to defend a theory of the phylogenesis of distinctively human mindreading (as dependent on distinctively human mindshaping), I need to say something about what is distinctive of human mindreading.

Many theorists of human social cognition have distinguished between high-level and low-level mindreading (Carruthers, 2006, 2009a; Apperly & Butterfill, 2009; Apperly, 2011), and this distinction comes close to the distinction I want to make between human and nonhuman mindreading. High-level mindreading typically involves the capacity to represent mental states *as such*. Most research into human social cognition has focused on the attribution of a specific category of mental states, that is, the propositional attitudes, like belief and desire; for this reason, I restrict my discussion to the propositional attitudes. Philosophers typically understand propositional attitudes and other mental states as concrete, unobservable causes of behavior. In addition, propositional attitudes are mental states with semantic properties: they represent the world as being a certain way, and this representation can be either satisfied or not, for example, in the case of beliefs, true or false, and, in the case of desires, fulfilled or unfulfilled. This is why the capacity to attribute false beliefs has figured so prominently as a test of full-blown human sociocognitive competence (Wellman et al., 2001). Furthermore, consensus has it that propositional attitudes can involve individually variable ways of representing the same facts, often called "modes of presentation." For example, Lois Lane can form beliefs about Superman by representing him not as Superman but as Clark Kent instead. Finally, it is widely held that propositional attitudes have tenuous connections to observable circumstances and behavior because of holism: how one reacts to a certain environmental situation depends not just on one propositional attitude but on indefinitely large networks of them (Morton, 1996, 2003; Bermúdez, 2003b, 2009). For example, the belief that it is raining alone does not trigger umbrella retrieval; it must be joined with a desire to stay dry that is stronger than competing desires, appropriate beliefs about locations of umbrellas, appropriate beliefs about the costs of umbrella retrieval, and so forth. Thus if high-level mindreading requires representing mental states as such, then high-level attribution of propositional attitudes requires representing them as unobservable, concrete causes of behavior, that (mis)represent the world as being a certain way, under individually variable modes of presentation, with complex connections to other propositional attitudes, perceptions, and behavior.

Apperly (2011) calls this characterization of high-level mindreading the "normative account": it specifies the competence attributors of propositional attitudes are *supposed* to have. However, as Apperly (2011) repeatedly illustrates, it is possible to pass behavioral tests of socio-cognitive competence with a less sophisticated understanding of propositional attitudes. For example, he argues that the standard test for whether or not human children can attribute false beliefs can be passed with no understanding that different agents might represent the same facts under different modes of presentation, or that propositional attitudes must combine with indefinitely many other mental states to yield behavior. According to Apperly (2011), infants and non-human animals often pass behavioral tests of propositional attitude understanding without fulfilling the normative account, i.e., without representing propositional attitudes as such. This is what he calls "low-level" mindreading. Apperly assumes that "low-level" mindreading still involves the attribution of unobservable causes that can misrepresent the state of the world or otherwise go unsatisfied. However, it does *not* involve an appreciation that the same situations can be represented under different modes of presentation, or that the link between observable situations and behavior is holistically constrained, i.e., mediated by indefinitely large networks of propositional attitudes.

Evidence suggests that some nonhuman animals, both closely and distantly related to *Homo sapiens*, are sensitive to or can track each other's propositional attitudes without necessarily representing them as such. That is, they can differentiate between behaviors caused by different beliefs and other mental states and respond flexibly and adaptively to these differences without thinking of these behaviors as caused by full-blown propositional attitudes. For example, chimpanzees appear to take into account whether or not a dominant conspecific has seen food being cached: if the dominant sees the caching, subordinates are less likely to access the cache (Hare et al., 2000, 2001). Western scrub jays, members of the amazingly precocious corvid family of birds, seem even more adept at this kind of mindreading than chimpanzees. Experiments have shown that these birds are much more likely to cache food in hard-to-observe locations, like behind barriers, in the shade, or farther away, when other birds observe them (Clayton et al., 2007). Given the intense competition for food among these birds, this is clearly a form of deception and requires some sensitivity to what conspecifics are likely to see. Such an interpretation is bolstered by evidence that western scrub jays keep track of what *different individual* conspecifics have witnessed when *recaching* food, and only individuals who have *themselves* pilfered others' caches engage in such deceptive strategies.

Whether or not such deceptive strategies require concepts of other minds and mental states, like false beliefs, is still the subject of much controversy. Even the most sanguine comparative psychologists restrict chimpanzee theory of mind to *some kind of appreciation* of conspecifics' goals, perceptions, and knowledge, denying that they appreciate false belief (Call & Tomasello, 2008). For example, sensitivity to a dominant conspecific's knowledge or ignorance of the location of recently cached food need not imply sensitivity to false beliefs. To track another's false belief, one must be capable of anticipating specific behaviors guided by the belief. But chimpanzees appear to know only that an ignorant, dominant conspecific is less likely to contest food retrieval, *not* the specific behaviors to which his false belief will lead, for example, searching where he thinks the food is (Hare et al., 2001). The jury, however, is still out. A definitive verdict on whether or not chimpanzees are sensitive to the contents of each other's beliefs must await new experimental approaches (Lurz, 2011). Anecdotal evidence suggests that chimpanzees can be extremely clever at tactical deception, and it is hard to explain such capacities without appeal to some understanding of false belief (Menzel, 1974). Nevertheless, even if a capacity for such *sensitivity* to false beliefs can be established, it would still not constitute high-level propositional attitude attribution in the sense I intend. As Apperly's (2011) thorough review of the comparative and developmental literature shows, an agent can be sensitive to false beliefs and other propositional attitudes without representing them as such.

The distinction between distinctively human and nonhuman mindreading therefore appears to be the following. Though nonhuman mindreaders show a kind of sensitivity to or ability to track at least some propositional attitudes, in their flexible and adaptive responses to the behaviors they cause, we find little evidence that this sensitivity is mediated by representations of propositional attitudes as such. Their sociocognitive feats do not require understanding that others' behavior is the product of unobservable *causes*, which represent situations under individually variable *modes of presentation* and influence behavior only tenuously, via *interaction with indefinitely large networks of other mental states* (Bermúdez, 2009). Although much human social cognition is plausibly similarly low level (Hutto, 2008), we are also capable of representing each other's propositional attitudes as such. The focus of this book is the evolution of the latter capacity: how and why did our species, and apparently *only* our species, evolve the capacity to understand behavior as caused by unobservable mental states, which represent situations under individually variable modes of presentation and influence behavior only tenuously due to holism, that is, constraint by whole networks of other mental states?

To make this question more precise, I contrast this characterization of full-blown, distinctively human mindreading with what Dennett (1987) has called adopting the "intentional stance." Dennett's position is a plausible characterization of a variety of low-level mindreading, of which some nonhuman animals and human infants are capable, and all normal humans employ in their unreflective, quotidian interactions. This is, admittedly, a departure from the letter of Dennett's discussions of the intentional stance; he often characterizes this notion as an analysis of full-blown propositional attitude attribution. However, Dennett's understanding of what propositional attitudes are is somewhat heterodox. He does *not* see them as concrete, unobservable mental states with causal control over behavior.[7] Rather, he sees them as abstract posits, akin to centers of gravity in physics, which help track robust patterns of observable behavior (Dennett, 1991b). In Dennett's view, to attribute propositional attitudes is *not* to speculate about the concrete causes responsible for behavior. Rather, it is to situate behavior in a *rational*, *normative* framework, to see it as a reasonable response relative to goals and available information (Zawidzki, 2012).

For example, in one of his most famous illustrations of the intentional stance, Dennett considers our interpretation of moves by a chess-playing computer (1978, pp. 4–9). According to Dennett, to interpret such moves from the intentional stance is *precisely not* to speculate about the algorithms that are causally responsible for them. We need know nothing about the design of the computer. Instead, interpretation requires only an understanding of *chess*: the goal of the game (checkmate) and the available information (piece configurations on the board, rules of chess, effective strategies). Applying this to quotidian interpretation of and by biological agents, adopting the intentional stance requires only the interpretation of bouts of behavior as goal directed and rationally constrained by available information, not the attribution of concrete, unobservable causes with content represented via individually variable modes of presentation.[8]

Both developmental (Gergely & Csibra, 2003) and comparative (Wood & Hauser, 2008) psychologists employ a similar framework for explaining low-level mindreading. According to Gergely and Csibra, infants as young as six and one-half months (Csibra, 2008) assume that agents pursue goals by the most efficient means available, given situational constraints. Wood and Hauser (2008) find evidence of similar reasoning in nonhuman primates. Such interpretive competence does not require speculating about concrete, unobservable causes of behavior or appreciating that these causes are full-blown propositional attitudes, that is, states with content represented via individually variable modes of presentation and holistically constrained influence on behavior. It requires only a sensitivity to certain abstract properties of bouts of behavior, namely, that they aim at specific goals and constitute the most rational means to those goals given environmental constraints. Gergely and Csibra (2003) call this the "teleological stance" or "the naive theory of rational action."

This very basic sociocognitive competence can also be supplemented with sensitivity to behavioral indicators of information access, like gaze direction, to yield an even more powerful understanding of rational action, which still falls short of high-level mindreading because it requires no attribution of concrete, unobservable mental states with content represented via individually variable modes of presentation and holistically constrained causal influence on behavior. Adopting this "enhanced teleological stance" (Zawidzki, 2011) allows interpreters to appreciate that different agents might have access to different information, and hence their goal-directed behavior might be rationally constrained by different situational factors. For example, subordinate chimpanzees that notice that dominant competitors have no line of sight on a location in which food is being cached apparently do *not* expect them to select means to the goal of food retrieval that are most efficient *relative to situational constraints of which only the subordinates are informed*. They take into account that different individuals are governed by different situational constraints, depending on the information to which they have access.[9] This "enhanced teleological stance" amounts to Dennett's intentional stance: behavior is predicted based on a rationality assumption governing the relationship between its goals and available information. Agents are assumed to engage in behavior that constitutes the most efficient means to goals relative to information to which they have access. If this is a tacit theory employed by human and nonhuman interpreters, it is a theory of observable behavior, not of the underlying mental causes responsible for it.[10]

One way of appreciating this, which plays an important role in the arguments of chapter 7, is that full-blown propositional attitude attribution supports, while adopting the intentional stance does *not* support, an appearance–reality distinction applied to agent behavior. The holism of the propositional attitudes ensures that any observed behavior is always compatible with mutually inconsistent propositional attitude attributions. Two qualitatively identical, counterfactually robust patterns of observable behavior may be caused by different propositional attitudes. This is not the case for adopting the intentional stance: if a particular attribution of goals and information access successfully rationalizes a pattern of behavior and supports prediction of future behavior, then, according to Dennett (1991b), there is no further fact of the matter about the agent's "true," unobservable beliefs and desires. Even if some pattern of behavior is indeterminate between multiple intentional stance interpretations, Dennett argues that there is no further fact of the matter about what propositional attitudes actually cause it; brute indeterminacy must simply be accepted. For this reason, most philosophers find Dennett's proposals hard to accept: it is difficult to give up the intuition that some determinate mental fact of the matter lies behind the behavioral appearances.

This characterization of low-level mindreading is even more deflationary than Apperly's (2011): not only are apparently sophisticated sociocognitive feats possible without an appreciation of different modes of presentation and the holistic connection between propositional attitudes and behavior, but there is no need to think of behavior as caused by concrete, unobservable mental states that can misrepresent the world. All that is necessary is a capacity to adopt the intentional stance, by which I mean a capacity to parse bouts of behavior into goals and rationally constrained means of achieving them, given information to which interpretive targets have access, where this access is understood entirely in terms of behavioral indicators, like gaze direction, and not in terms of unobservable cognitive states, like beliefs.

Thus I assume the following taxonomy of varieties of mindreading. The social cognition of nonhuman animals, human infants, and human adults engaged in unreflective, quotidian interactions is often guided by tacit knowledge of behavioral patterns, sometimes highly abstract ones, involving categories like "goal," "efficient means," "information access," and "teleological rationality." The most sophisticated examples of such low-level mindreading plausibly involve adopting something like Dennett's intentional stance. As Dennett (1991c) himself makes clear, this is better characterized as an unreflective, tacitly encoded "craft" than an explicit theory. Low-level mindreading can, of course, involve even less sophisticated representations of behavioral patterns. For example, nonhuman animals and human infants use straightforward induction to anticipate future behavior and are sensitive to various nonrational behavioral regularities, such as correlations between facial expressions of emotion and subsequent behavior (Parr, 2001; Andrews, 2007, 2008).

In addition, some adult human social cognition is guided by the representation of propositional attitudes as such. We can predict each other's behavior based on attributions of concrete, unobservable mental causes, which represent situations under individually variable modes of presentation and must combine with indefinitely broad networks of other mental states to yield behavior. This is more than a tacit theory of observable behavior; it is a theory of the underlying mental causes of observable behavior. One of the main goals of the book is to show that the evolution of this latter form of sophisticated, high-level mindreading depended on already extant, sophisticated, and distinctively human mindshaping practices. As will become clear, it is likely that such mindshaping practices both required and selected for more sophisticated versions of the intentional stance than those of which contemporary nonhuman animals are capable. However, these mindshaping practices did not require high-level mindreading in the sense defined here, and such high-level mindreading was not possible without them.

4 Mindshaping

The term "mindshaping" is not commonly used in the literature on human social cognition. I adopt it from Mameli (2001). Mameli argues that through the mechanism of social expectancies, folk assumptions about human psychology can become self-fulfilling prophecies. For example, assumptions about gender lead adults to expect male and female infants to have different psychologies: while boys are supposed to be aggressive and quick to anger, girls are supposed to be passive and easily upset. For this reason, adults tend to interpret the same behavior differently depending on perceived gender. If they think a crying infant is a boy, they interpret him as being angry. If they think the same crying infant is a girl, they interpret her as being upset. These interpretations give rise to social expectancies: adults expect a crying male infant to act in aggressive ways and a crying female infant to act in passive ways, for example, to seek comfort. Such expectancies affect the way adults interact with infants, for example, being more quick to comfort a crying female infant than a crying male infant. Such differences in patterns of interaction might lead to self-fulfilling prophecies: infants come to behave in ways that are consistent with adult expectancies. Mameli argues that a similar dynamic might be at work when parents interpret early infant behavior as intentional communicative acts: infants begin to act in ways that confirm such interpretations, and this helps "bootstrap" the capacity for intentional communication in human infancy (2001, pp. 617–619). In both of these cases, argues Mameli, folk psychological behavioral interpretation functions to shape infant minds rather than read them.

Mameli sees mindshaping as a kind of “niche construction” (Laland et al., 2001). Niche construction is a relatively new concept in biology. The idea is that evolution does not always consist in adapting to a prespecified environmental niche through genetic selection. Often species construct environmental niches that are a better fit for their current genes. This might involve nothing more elaborate than imprinting on a new source of food or shelter. For example, Mameli considers the hypothetical case of a species of butterfly that imprints on the plant on which it feeds after hatching. That is, mature butterflies tend to lay eggs on plants of the same species as those on which they hatched, recognizing them via some signal. Imagine that this imprinting mechanism is fallible: sometimes mature butterflies lay their eggs on the wrong species of plant. Once, when this occurs, the deviant plant species fortuitously happens to be a better source of nutrition for that species of butterfly. The lucky butterfly has more descendants than those of its conspecifics that do not mistakenly lay eggs on the plant. These descendants imprint on the new plant, and the species’ niche has been altered without any genetic selection: a new selection pressure arises because butterflies that lay eggs on the new plant do better than butterflies that do not make this “mistake.” Imprinting is a form of nongenetic trait inheritance that can alter a species’ niche in ways that feed back into genetic inheritance. Mameli’s idea is that mindshaping via the mechanism of social expectancies is a human form of niche construction. We alter the selectional environment of subsequent generations by shaping their minds in ways that affect the social niche in which they find themselves.

Sterelny (2012) argues that distinctively human social cognition drove such social niche construction in human evolution.[11] He considers two different models of the evolution of social cognition in the human lineage. In both views, the central problem that human social cognition must solve is cooperation. In the received view—the Machiavellian intelligence hypothesis—the problem of cooperation can be solved only by detecting and preventing “free riders” from taking advantage of the cooperative efforts of others. But this sets up an arms race pitting better deception against better deception detection. The result is runaway selection for better mindreading. Suppose, for example, that a mutant with better mindreading ability is introduced into a prehuman population. The mutant uses her skill to successfully deceive her conspecifics, accumulating more survival-related resources, living longer, and having more offspring. Soon better mindreaders dominate the population. But since hominid biological success depends crucially on social interactions, this amounts to a dramatic alteration of the hominid niche. Where previously it was possible to succeed without mindreading virtuosity, this is now impossible. There are now strong selection pressures favoring good mindreaders. The mutant mindreader has not just adapted to a preexisting niche. She and her descendants have significantly altered the social niche to which subsequent generations must now adapt.

Sterelny favors a different model of the interaction between human social cognition and social niche construction. In his view, the key sociocognitive innovations in the human lineage are capacities for high-fidelity transmission, preservation, and elaboration of information across generations. As with Machiavellian intelligence, such capacities are plausible responses to the problem of cooperation, though to a different aspect of this problem. Sterelny argues that detecting and dissuading free riders was not an important problem among small bands of social foragers early in our lineage. Coordination was more important: advanced planning and on-the-spot adjustment to prey were necessary to successfully hunt the large fauna that early humans and their immediate precursors hunted with their relatively crude weapons. But such coordination required effective communication and training, possible only through high-fidelity cultural learning. Cooperative hunts yielded such nutritional boons that strong selection pressures favored a capacity to transmit and preserve information through cultural learning. This dramatically altered the social niche for subsequent generations. Individuals who were not good cultural learners could not benefit from group hunts and were at a major selectional disadvantage.

As I argue in chapter 4, in examining the evolution of distinctively human mindshaping, I think that Sterelny's picture of social niche construction in the human lineage is closer to the truth than the picture suggested by the Machiavellian intelligence hypothesis. For now, however, the important lesson is that evolutionary change in hominid social cognition can have dramatic feedback on hominid evolution: it results in an alteration of arguably the most important component of the hominid niche, that is, social circumstances, thereby creating new selection pressures. Mindshaping plays an important role in this process. If individuals are rewarded for conforming to social expectancies of the kind discussed by Mameli, or other kinds related to the cooperative projects highlighted by Sterelny (e.g., coordination and communication conventions passed down through cultural learning), then this constitutes the construction of a new social niche to which subsequent generations must adapt. Such niche construction can thereby result in selection for new genes, for example, encoding more cooperative behavioral dispositions.

As I intend to use the term, "mindshaping" refers to the kinds of phenomena discussed by Mameli, together with numerous other human social behaviors. Here are some examples: distinctively human imitation, pedagogy, normative judgment and norm enforcement, the institution of social roles, and self-constituting narratives. An important goal of this book, especially chapter 2, is to show that such behaviors, often treated as unrelated, all share important properties related to mindshaping and niche construction. All such human practices aim to get a target to match the behavior of some model, and this, I argue, is the essence of mindshaping. Nonhuman species show only the rudiments of some kinds of mindshaping, for example, a limited capacity to imitate. On the other hand, human beings are obsessive mindshapers. It is now well established by studies of very young infants that human beings are wired to be receptive to pedagogical instruction (Csibra & Gergely, 2009) and “overimitate” (Nielsen & Tomaselli, 2010) models, eagerly acquiring even noninstrumental behaviors, that is, behaviors that are not essential to securing typical goals like acquiring a favored food. We seem to engage in mindshaping, both as shapees and as shapers, for its own sake. As I argue in chapter 2, when it comes to social behavior, our species is best distinguished from others by the complexity, subtlety, variety, and broad scope of human mindshaping, and the inordinate amount of time and resources we devote to it. Chapter 4 addresses the evolution of such mindshaping, arguing that it constitutes a kind of *targeted* social niche construction, making successful human social interaction, on cooperative projects especially, significantly more computationally tractable than it would be otherwise.

Thus although Mameli’s notion of mindshaping is my point of departure, I defend an expanded version of the concept. It applies to a far greater variety of human practices than Mameli (2001) envisions. Furthermore, although I agree with Mameli that mindshaping functions as a method of social niche construction, I emphasize that distinctively human mindshaping makes possible a kind of social niche construction that is qualitatively distinct from those available to other species. In Mameli’s example of the butterfly accidentally imprinting on a more nutritious plant, the niche construction is a fortuitous by-product of traits selected for effects that have nothing to do with niche construction. Butterfly imprinting on plant species is selected for its nutritional effects on offspring. It then indirectly affects niche. But mindshaping practices like imitation, pedagogy, norm enforcement, the institution of social roles, and narrative self-regulation are directly targeted at social niche construction. This is not to say that practitioners consciously conceive of themselves as constructing social niches for subsequent generations. The point is, rather, that such behavioral traits are selected for their effects on the social niche, making populations more cooperative, homogeneous, and predictable. So, I argue, distinctively human mindshaping is crucial to explaining the success of the hominid sociocognitive syndrome because it constitutes a way of bringing social niche construction under control: unlike other species, we obsessively engage in practices whose raison d'être is social niche construction. Unlike fortuitous niche construction that occurs as a by-product of traits selected for other reasons, human mindshaping enables *targeted* social niche construction.[12] This is key to understanding mindshaping's crucial role in the evolution of the human sociocognitive syndrome.

5 Cooperation

Compared to other primates and, indeed, most other vertebrates, humans are an uncommonly cooperative species (Sterelny, 2003, 2007, 2012). There is no doubt that cooperation is central to our relative biological success. We depend on cooperation to secure all our biologically significant goals: feeding, defense against predation and other threats, mating, reproduction, and care of offspring. Although some nonhuman primate species engage in rudimentary forms of cooperation, like monkey hunting among some chimpanzee populations (Boesch, 1994), nothing comes close to matching the breadth, scope, sophistication, pervasiveness, and centrality of cooperative projects among humans. Cooperation is so fundamental to the human way of life that we often fail to notice how much depends on it; we take it for granted like the proverbial fish in water. For example, sophisticated communication is possible only against a background of cooperation. Language is impossible without truth telling and trust, requiring cooperative tendencies unmatched by other hominids. Overwhelming evidence indicates that the disposition to cooperate is a universal human trait. Cross-cultural experiments involving economic games played for real money show that human beings of all cultures tend to favor equitable distributions of goods (Henrich et al., 2006). Developmental psychologists have shown that even very young children are default cooperators (Tomasello, 2009). Few claims are as well established in the social sciences as the claim that pervasive cooperation is a universal and distinctively human trait.

Thus it is fair to conclude that cooperation is one of the central functions of human social cognition. We need to be good mindreaders, mindshapers, and communicators because these proficiencies are necessary for successful cooperation, the key to our biological success. This basic premise, forcefully defended in recent years by Kim Sterelny (2003, 2007, 2012), looms large in the arguments of this book. In particular, chapter 4 argues that human cooperation cannot be explained by sophisticated mindreading. For this reason, since cooperation is one of the central functions of human social cognition, mindreading is not its most important component. Instead, I argue in chapter 4 that sophisticated mindshaping is the key to understanding how human cooperation is possible.

Despite its clear importance to the human species, the evolution of cooperation is notoriously difficult to explain. The reason is simple: it is easy to take advantage of overly cooperative agents. Imagine some protohuman hominid population no more cooperative than contemporary chimpanzees. Next imagine introducing a mutant into this population who is more cooperative. Perhaps the mutant shares food spontaneously. It is clear that such a mutant is unlikely to leave many offspring relative to her conspecifics. Other things being equal, the mutant will secure fewer net calories than her conspecifics, leaving her at a competitive disadvantage when it comes to securing mates, caring for offspring, and defending against predation. So it is hard to see how the kind of pervasive, default cooperation that characterizes our species can ever get off the ground. Even if this could be explained, this would lead to another puzzle: how can such cooperation remain stable in a population? Imagine that a mutant is introduced into a population of default cooperators. The mutant is a "free rider": she takes advantage of the fruits of others' cooperative efforts without doing her share. She secures all her caloric needs without expending the energy or incurring the risk associated with hunting or foraging. She takes advantage of offspring care offered by her cooperative conspecifics without reciprocating. Other things being equal, such a mutant would have more offspring than any of her conspecifics, and after several generations, her descendants would dominate the population.

Researchers have proposed a number of well-known solutions to these puzzles. The most well-confirmed model appeals to inclusive fitness or kin selection (Hamilton, 1964). It is worth sacrificing one's own biological prospects to cooperate with genetically related conspecifics. This occurs because selectively helping those who carry one's own genes helps ensure that those genes are passed on, including the genes for helping one's genetic relatives. Such kin selection explains most cooperation observed among nonhuman species. It is why many species of insects include classes of individuals willing to sacrifice their lives for the "queen" (Cronin, 1991). It is also why mammalian mothers go to extraordinary lengths to provision and protect their offspring. However, it is obviously insufficient to account for human cooperation, which involves trusting and forgoing advantages for genetically unrelated, and often completely unfamiliar, individuals.

Proposed solutions to the puzzle of cooperation with genetically unrelated individuals appeal to various kinds of reciprocity. The simplest variety, direct reciprocity, involves keeping track of whether or not an individual has cooperated in previous interactions. Perhaps the most famous example of this kind of reciprocity comes from computer simulations of interacting agents (Axelrod, 1984). Simulated agents would play the "prisoner's dilemma"[13] game against each other in tournaments, accumulating points, depending on their record of victories and defeats. Although in a one-shot prisoner's dilemma the uncooperative agent always comes out on top, this is not necessarily the case in iterated versions. One of the most successful simulated agents implemented the so-called "tit for tat" strategy: when interacting with another agent for the first time, "tit for tat" always cooperates, but on subsequent interactions with that agent, "tit for tat" does whatever that agent did in the previous interaction: cooperates if it cooperated, defects if it defected. Such computer simulations show that cooperation can be rewarding if agents interact repeatedly with each other. However, it is generally acknowledged that keeping track of numerous interactants' histories of cooperation is too cognitively demanding to explain cooperation among unrelated individuals in hominid populations (Stevens & Hauser, 2004; Henrich, 2004; Stevens et al., 2005). Also, subsequent computer simulations show that "tit for tat" does not do well against more sophisticated though less cooperative strategies (Dugatkin, 1997).

In response to such problems with simple direct-reciprocity models of cooperation, more complex versions of reciprocity have been proposed. For example, in "indirect reciprocity," the population of interactants also speaks a language that can be used to report on particular individuals' reputations for cooperation (Nowak & Sigmund, 2005). Cooperation is motivated by the long-term advantages of gaining a reputation for cooperation. Even if one never interacts with a particular individual again, or if one cannot remember whether or not the individual cooperated in previous encounters, it is worth cooperating because the individual with whom one cooperates will then communicate one's reputation as a cooperator to other members of the population, making them more likely to cooperate with the original agent in the future. The problem with this proposal is that communicating reputation is itself a cooperative act (Henrich, 2004). What is to stop one's interactants from spreading false rumors about one's cooperative inclinations? So-called strong reciprocity is another model proposed to avoid the problems raised by direct and indirect reciprocity (Henrich, 2004). Strong reciprocity is the disposition not just to cooperate but to punish, at a cost to oneself, uncooperative behavior in others. This raises the costs of uncooperative behavior and reduces the relative costs of cooperative behavior (Sigmund, 2007). However, like indirect reciprocity, strong reciprocity presupposes a form of cooperation that it cannot therefore explain. The reason is that costly punishment is itself a cooperative act: one incurs the cost of punishing the uncooperative, thereby making life easier for cooperative individuals who do not punish (Henrich, 2004). Such "second-order free riders" would eventually drive strong reciprocators to extinction.

Thus the remarkable cooperativeness of human beings remains a recalcitrant puzzle. The various proposed solutions reviewed here have been significantly tweaked in recent years to avoid the obvious problems I have briefly reviewed. Chapter 4 looks at such models in greater detail and argues that ultimately, dispositions to shape minds and have our minds shaped to respect prosocial norms are central to explaining the evolution of human cooperation.

6 Complex Communication

The final component of the human sociocognitive syndrome is complex communication. Public language is, of course, the best example of this; however, there are others, for example, music, ritual, and dance. Other species communicate, and sometimes their communicative behaviors can be highly complex, for example, birdsong. But human communication stands out because it marries extremely complex structure with extremely flexible use. For example, the recursive syntax of human language can generate well-formed formulas of arbitrary complexity. At the same time, such arbitrarily complex structures express a generally systematic semantics. We can use language to encode messages about almost anything, including facts that are not perceptually salient, or indeed that could not be perceptually salient, like subatomic structure. Such general expressive capacity is nonetheless systematic. There is a finite set of strict rules for pairing semantics with expressions, allowing for the well-ordered construction of an infinite variety of messages.

These rules permit the combination of arbitrarily diverse contents into unified expressions. For example, human language meets Gareth Evans's "generality constraint" (1982): one can combine any subject with any predicate. This is why we can communicate metaphorical messages, like "Light is a wave," which play an important role not just in poetry and literature but in science as well. Even the structurally most sophisticated examples of nonhuman communication, like birdsong, are extremely impoverished in the variety of messages they can convey. Birdsong, for example, is used almost exclusively to advertise male fitness in the context of courtship displays (Fitch, 2004).

The attempt to explain how a communicative system as structurally complex and semantically flexible as human language evolved has had a notoriously controversial history. Few evolutionary puzzles have generated as great a diversity of underconfirmed just-so stories. The received view sees language as a tool for externalizing thought (Pinker & Bloom, 1990). According to this view, the thought of our nonlinguistic precursors already had the structural complexity and semantic flexibility of human language. At the same time, sharing such thought had potentially dramatic, positive effects on fitness. So human language evolved as an adaptation for sharing thought, inheriting its structural complexity and semantic flexibility from the thought it evolved to express. This intuitively compelling picture fits well with the received "mindreading as sociocognitive linchpin" theory. Language is seen as a tool selected primarily for enhanced mindreading: for helping individuals learn each other's thoughts, where these are understood as constituted independently of the linguistic means used to express them. Furthermore, language use presupposes already extremely sophisticated mindreading abilities, especially the capacity to recognize communicative intentions, including which belief a speaker intends to express. To put it succinctly, according to the orthodox understanding of language evolution, language evolved as a way of enhancing the kind of sophisticated mindreading that made it possible in the first place (Origgi & Sperber, 2000).

A variety of problems undercut this intuitively very compelling picture. Most obviously, it presupposes a solution to the problem of cooperation. Sharing one's thoughts is a cooperative act. An honest communicator shares information that can be used against her without receiving anything in return. How can a disposition for honest communication win out against deceptive strategies? Another, related problem is a version of the holism problem that arises for all forms of sophisticated mindreading. If interpreting another's utterances requires inferring her communicative intentions and, in particular, the beliefs she wants to express, how is this possible if any belief is compatible with any behavioral evidence, given appropriate modifications of background propositional attitudes? If another's utterances are supposed to be evidence of her beliefs, thereby enhancing mindreading, how is the problem of uncooperative or otherwise inappropriate background intentions, for example, intentions to deceive, mitigated?

Finally, evidence strongly suggests that the thought of our closest nonhuman relatives is not structurally complex and semantically flexible in the way that language is. The kinds of thoughts they have to communicate do not require a language as powerful as human language. There is reason, therefore, to conclude that our last common ancestor with them also lacked thought with the structure and semantic flexibility of human language. And it is difficult to imagine how such thought can have evolved between the time of the LCA and the origin of our own species, with, presumably, the capacity for acquiring human language. So how can this capacity have evolved as a tool for expressing such thought? In fact, as Chomsky and his heirs point out, most thoughts that humans need to communicate to gain biological advantage do not require such a complex and semantically flexible system of communication. Speakers of pidgins and other languages constructed on the fly by interactants who do not share a public language (what Klein & Perdue [1997] call the "basic variety") do very well in their pragmatic collaborations. So if language was selected for the communication of biologically significant information, why all the apparently excess capacity? If the thought of our prelinguistic precursors, like that of contemporary chimpanzees, lacked the structure and semantic flexibility of contemporary language, then why did a communicative system as structurally complex and semantically flexible as human language evolve?

As I argue in chapter 5, the key to resolving these puzzles is conceiving of language as a mindshaping device, rather than as a mindreading device. I stress the continuity between language and other structurally complex yet semantically less precise forms of human communication, like music, dance, and ritual. We have increasingly compelling evidence that music, dance, and ritual can play an extremely direct role in enhancing cooperation (Wiltermuth & Heath, 2009; Kirschner & Tomasello, 2010). They also play an obvious role in distinguishing group members from nonmembers. It is widely assumed that such between-group distinctions play an indispensable role in making group selection possible, and group selection is key to understanding the evolution of human cooperation. Music, dance, and ritual can function to increase within-group homogeneity and between-group differences. They do so by shaping the minds of participants to be more alike. Furthermore, such forms of communication do not presuppose the kind of structurally complex, flexible thought that mirrors contemporary human language.

Accordingly, music, dance, and ritual are ideally suited to play the role of a kind of "missing link" between contemporary human language and nonhuman communication systems. The structural analogies between music and other rhythmic behaviors and contemporary language are significant (Lerdahl & Jackendoff, 1996). Such affinities have led some to propose that song and dance are indeed precursors to full-blown human language (Darwin, 1871/1981; Okanoya, 2002; Mithen, 2006; Fitch, 2010). So a musical protolanguage, employed in ritual, can show how creatures incapable of thought that mirrors the structural complexity and semantic flexibility of contemporary language could nonetheless use complex communicative behaviors to shape each other in ways that enhanced cooperation and mutual interpretability. We can then conceive of contemporary language as a highly refined descendant of such early mindshaping, cooperation- and interpretation-enhancing communicative systems. For example, whereas our prelinguistic precursors could use only relatively imprecise rhythmic rituals to commit to vaguely defined social roles, we are now able to use language to commit to courses of behavior of arbitrarily precise specificity, for example, the role of a believer that *p*, where *p* stands for any declarative sentence expressible in language. We can then use language to share and read each other's thoughts because we use language to make commitments to behavior that constitute such thoughts.

This, in any case, is the picture I defend in chapters 5 and 7. Language is key to understanding sophisticated mindreading *not* because it evolved as a way of externalizing independently constituted thoughts that share its structure and semantics, like the propositional attitudes; rather, language makes possible sophisticated mindreading because it helps constitute such thoughts by enabling us to commit to behavior consistent with claims we make.

7 Two Pictures of the Evolution of Human Social Cognition

The view that I criticize is intuitively compelling. It is hard to explain how human beings manage to anticipate each other's behavior with the precision and reliability necessary to explain our cooperative and communicative feats, without assuming that we are extremely proficient mindreaders. In this picture, the first and most important evolutionary step in the direction of the human sociocognitive syndrome was the emergence of a population of natural psychologists of unparalleled skill. Before we could shape each other to be more cooperative, before we could communicate using a language of unmatched syntactic complexity and semantic flexibility, we must have been able to reliably infer, based on behavioral evidence, the propositional attitudes responsible for each other's behavior. After all, how can you shape people to be more cooperative and more communicative without first understanding what and how they think?

Despite its intuitive appeal, I think this picture is fundamentally wrong. Inferring another's propositional attitudes based on her behavior is a computationally intractable task, unless she has already been shaped to be cooperative and easily interpretable. Such shaping does not require prior mastery of human psychology. As with other evolved capacities, gradual and piecemeal accumulation of dispositions to mindshape that happened to work in the contingent ecological contexts in which our ancestors found themselves gave rise to more humanlike cooperative and communicative practices. Such gradual development did not require deep psychological insights. Instead, motivations to mindshape coupled with modestly enhanced versions of the intentional stance employed by contemporary chimpanzees were sufficient to give rise to complex cooperation and communication. These then made possible sophisticated mindreading by providing culturally shared frameworks that interactants used to shape their thoughts and behavior to be easily interpretable. That, in any case, is what I seek to establish in the chapters ahead.