Chapter 1
Mindblindness and Mindreading

Imagine what your world would be like if you were aware of physical things but were blind to the existence of mental things. I mean, of course, blind to things like thoughts, beliefs, knowledge, desires, and intentions, which for most of us self-evidently underlie behavior. Stretch your imagination to consider what sense you could make of human action (or, for that matter, any animate action whatsoever) if, as for a behaviorist, a mentalistic explanation was forever beyond your limits. This is a hard thought experiment. See if it helps to make it more concrete by considering how we understand even a simple human act:

> John walked into the bedroom, walked around, and walked out.

To make sense of this, we ask ourselves why John behaved in this way. A mindreader might answer this question by saying something like this:

> Maybe John was **looking** for something he **wanted** to find, and he **thought** it was in the bedroom.

Or the mindreader might think:

> Maybe John **heard** something in the bedroom, and **wanted to know** what had made the noise.

Or:

> Maybe John **forgot** where he was going: maybe he really **intended** to go downstairs.

A mindreader can generate a longish list of such "maybes" to explain John's behavior—and it is a safe bet that most of them will be based on John's mental states. (In the examples above, the mental-state words are printed in boldface to make it easy to pick them out.)

Now, you and I are mindreaders. I don't mean that we have any special telepathy (see Whiten 1991); I just mean that we have the capacity to imagine or represent states of mind that we or others might hold. Mindreading is nothing mysterious; however, as I hope to show in this book, it is impressive.

Notice that in the above examples our way of thinking about mental states is prefixed by "maybe." We are never 100 percent sure what we or others are thinking (since mental states are to some extent hidden from view), but we nevertheless find it easy to imagine what others may be thinking.

What sense does a person with mindblindness make of John's behavior in the example above? In trying to answer this, we must of course refrain from using any mental-state terms in the explanation. Here is an attempt:

> Maybe John just does this every day, at this time: he just walks into the bedroom, walks around, and walks out again.

Notice that this is not an explanation in terms of any causal motive or reason. Rather, it is simply a statement about possible temporal regularities. It is also very likely to be wrong. When our mindblind person discovers that John does not do this every day at this particular time, he or she will need to come up with another non-mentalistic attempt to explain John's action.

The problem is that there just aren't many simple, readily available, plausible, non-mentalistic explanations for John's behavior. (Try generating some yourself if you don't believe me.) To a person with mindblindness, even this very basic sequence of acts—walking into the bedroom, walking around, and then walking out again—is a real mystery. Now imagine what sense a mindblind person would make of an infinitely more complex social situation (one I observed in my local park in Islington):

> Joe and Tim watched the children in the playground. Without saying a word, Joe nudged Tim and looked across at the little girl playing in the sandpit. Then he looked back at Tim and smiled. Tim nodded, and the two of them started off toward the girl in the sandpit.

Again, as mindreaders, we make sense of the situation in mentalistic terms right from the start. For example, we might come up with a rather sinister reading of the situation:

> Maybe Joe and Tim had a **plan** to do something nasty to one of the children. Joe **wanted** Tim to **know** that their victim was to be the little girl in the sandpit, and he **indicated** this by the direction of his gaze. Tim **recognized** Joe's **intention** and nodded to tell Joe he had **understood** the **plan**. Then they went over to the little girl, who was **unaware** of what was about to happen.

Or we might come up with a more rosy reading of the situation:

> Maybe Joe **wanted** to point out to Tim who it would be fun to play with. Tim **agreed** with Joe's **idea**, so they went over to ask the little girl in the sandpit if she **wanted** to play.

Note that both of these interpretations are littered with mental-state concepts and with terms that express these concepts. (Just look at the bold type.) Indeed, it is hard for us to make sense of behavior in any other way than via the mentalistic (or "intentional"[1]) framework. We just can't help doing it this way; as Fodor (1983) emphasizes, it is a consequence of our biology.

When someone points out all this mindreading to you, it hits you with some force. Recall the apocryphal man who was shocked to discover he had been speaking in prose all his life. We mindread all the time, effortlessly, automatically, and mostly unconsciously. That is, we are often not even aware we are doing it—until we stop to examine the words and concepts that we are using.

What sense would a person with mindblindness make of Joe and Tim in the scene described above? Precious little, is my guess. Why were Tim and Joe smiling at each other? And what did those glances mean? Why did they move off together in the direction of the little girl?

Without a mentalistic framework—or, as Dennett (1987) calls it, the Intentional Stance—a person with mindblindness is thrown back on temporal-regularity accounts or on routine-script explanations (like the one I came up with for why John walked into the bedroom) or is forced to use unwieldy things resembling the "reinforcement-schedule" explanations that behaviorist psychologists construct. None of these seem very useful here. The first two are too limited in their application to the constantly changing social world; the third takes too long to compute. In the heat of a social situation, it pays to be able to come up with a sensible interpretation of the causes of actions quickly if one is to survive to socialize for another day. Nonmentalistic explanations are just not up to the job of making sense of and predicting behavior rapidly. Instead, a person with mindblindness is left confused: Just what are Joe and Tim up to? In the meantime, the mindreader sizes up the situation instantly.

It is probably impossible to imagine what it is like to be mindblind, in the same way as it is impossible to imagine what it is to be a bat (Nagel 1974).[2] To live in a bat's world, in which objects are known by echo location, must impart a notion of objects so radically different from the notion that we obtain through vision that it may be beyond our imagination. Conversely, it is probably impossible for a mindblind person to imagine what it is like to be a mindreader. In the words of Sperber (1993), "attribution of mental states is to humans as echolocation is to the bat." It is our natural way of understanding the social environment.

The gulf between mindreaders and the mindblind must be vast. Gopnik (1993) gives a vivid account of her attempt to imagine just what the world must look like through the eyes of someone who is mindblind:

> This is what it's like to sit round the dinner table. At the top of my field of vision is a blurry edge of nose, in front are

> waving hands. . . . Around me bags of skin are draped over chairs, and stuffed into pieces of cloth, they shift and protrude in unexpected ways. . . . Two dark spots near the top of them swivel restlessly back and forth. A hole beneath the spots fills with food and from it comes a stream of noises. Imagine that the noisy skin-bags suddenly moved toward you, and their noises grew loud, and you had no idea why, no way of explaining them or predicting what they would do next.

Tragically, mindblindness is not an idle thought experiment or a piece of science fiction. For some people, it is very real. Gopnik hints at how terrifying it would be to be mindblind. I think she must be right—I certainly would not want to be without the ability to read behavior in terms of mental states. In this book I will discuss the idea that children and adults with the biological condition of autism suffer, to varying degrees, from mindblindess.[3] For reasons to be explored, they fail to develop the capacity to mindread in the normal way.

I intend in this book to extend the ideas first proposed by Nicholas Humphrey in a series of groundbreaking essays published in the 1970s and the 1980s and collected in Humphrey 1984. Humphrey had the brilliant insight that the best way to characterize humans is as *Homo psychologicus*. As he put it (Humphrey 1984, p. 3), "Human beings are born psychologists." He elaborated this as follows:

> The ability to do psychology, however much it may nowadays be an ability possessed by every ordinary man and woman, is by no means an ordinary ability. . . . The fact is that, whatever may be the logical problems of describing inner experience, human beings everywhere openly attempt it. There is, so far as I know, no language in the world which does not have what is deemed to be an appropriate vocabulary for talking about the objects of reflexive consciousness, and there are no people in the world who do not quickly learn to make free use of this vocabulary. Indeed, far from being something which baffles human

> understanding, the open discussion of one's inner experience is literally child's play to a human being, something which children begin to learn before they are more than two or three years old. And the fact that this common-sense vocabulary is acquired so easily suggests that this form of description is natural to human beings precisely because it maps directly onto an inner reality which each invidual, of himself, innately knows. (ibid., pp. 5, 8)

Humphrey's argument, then, is that the ability to see behavior in terms of an agent's mental states is inborn and is the result of a long evolution. Consequently, Humphrey concludes that at one point in our evolution there must have been a time when we lacked this ability. Here is the "Just So" story in which he first proposed this idea:

> . . . once upon a time there were animals ancestral to man who were not conscious. That is not to say that these animals lacked brains. They were no doubt percipient, intelligent, complexly motivated creatures, whose internal control mechanisms were in many respects the equals of our own. But it is to say that they had no way of looking in upon the mechanism. They had clever brains, but blank minds. Their brains would receive and process information from their sense-organs without their minds being conscious of any accompanying sensation, their brains would be moved by, say, hunger or fear without their minds being conscious of any accompanying emotion, their brains would undertake voluntary actions without their minds being conscious of any accompanying volition. . . . And so these ancestral animals went about their lives, deeply ignorant of an inner explanation for their own behavior. (ibid., pp. 48–49)

In calling this a "Just So" story I do not intend to belittle Humphrey's major contribution. It is his own description of his work. He did not attempt to prove the idea scientifically (indeed, there is a real question as to whether such evolutionary claims can ever be proven), so this is just about as far as he took this momentous idea. He of course added some speculations of his own, such as the idea that we mindread by using our own experience of introspection as a simulation of another's mental states—an idea that has been discussed by subsequent "simulation theorists."[4] But aside from these speculations, Humphrey has left it to others to take up his key idea and develop it scientifically.

I hope my own work will contribute to this enterprise. In this book I try to extend Humphrey's idea by presenting a model of the psychological development of mindreading, proposing some fundamental mechanisms that I think must underlie this remarkable ability. I then explore the evidence that children with autism fail to develop this capacity. Next I consider how this might throw light on the evolutionary and neurobiological bases of this capacity. I also argue that the study of autism highlights the role played by two mechanisms that allow us to understand the "language of the eyes."

Chapter 2
Evolutionary Psychology and Social Chess

The approach I take in this book can in part be summed up as an attempt to contribute to the study of evolutionary psychology. Cosmides, Tooby, and Barlow (1992, p. 7) define this as "psychology informed by the fact that the inherited architecture of the human mind is the product of the evolutionary process."

What Is Evolutionary Psychology?

"Evolutionary psychology," of course, strikes a chord with "evolutionary biology," which has transformed the science of biology. As Cosmides et al. suggest, the time is now ripe for psychology to be integrated with biology via evolutionary theory. Every branch of biology, from biochemistry and genetics to zoology and ecology, has been incorporated into a Darwinian framework, and it is something of an embarrassment that psychology has dawdled so far behind in this. One thing I am trying to do here is make up for this by sprinting forward to link arms with psychology's neighbor disciplines in biology.

I do not mean to suggest that all of psychology will be amenable to the Darwinian framework, since it is clear that there will always be a slice of human behavior that has nothing to do with natural selection. (Consider, for example, fashion in clothing.) But wherever psychologists are investigating human universals—the universal aspects of language, color vision, parenting, or (as I argue) mindreading, to name just a few—there is the strong likelihood that the phenomena are biological, innate, and products of natural selection.

A word of qualification is needed about the use of the word "universal." If something is universal, it does not mean that there is no variation in the phenomenon across individuals. On the contrary, it is highly likely that there will be variation, just as there is considerable variation in the size of human stomachs. The point is that every human body *has* a stomach! It is in this sense that I will be arguing that mindreading is universal. Of course a genetic pathology might result in the birth of an individual without a stomach, but such a genetic aberration would not undermine the notion that the stomach is universal; indeed it might help us identify the genes that control the building of the normal stomach. Paradoxically, the genetic pathology that caused a body to be built without a stomach would be further evidence for the evolution of the stomach.

A qualification about the use of the word "biological" is also needed. When I say that a psychological state is biological, I mean that there is a specific process in the brain that controls it. In a sense, all psychological states are biological—"no brain, no mind," goes an old adage. Here I am trying to highlight those psychological states (e.g., talking) that are inevitably universal, in contrast to those (e.g., watching television) that might be universal but are not inevitably so. The former are more like biological instincts—hence Pinker's (1994) use of the phrase "the language instinct." In this book, I will be concerned with what might be thought of as "the mindreading instinct."

Steven Pinker has fruitfully adopted the evolutionary approach to language, using exactly the logic mentioned earlier: try finding a human society in which the people have no language. Such societies are nonexistent, Pinker argues, precisely because language is part of human biology. Cultural variation, of course, is massively evident among the world's 6000 or so languages, but the basic drive to develop and use language—the language capacity—is universal. (Pinker, following Noam Chomsky, also summarizes the evidence for syntactic universals.)

Similarly, the evolutionary approach to parenting is a central part of Bowlby's (1969) "attachment theory."[1] It is clear that human infants (like the immature members of a long line of ancestor species) have a strong drive to "attach" to an adult caregiver, and this is highly adaptive both in terms of the infant's physical survival and in terms of its psychological well-being. Regarding the study of color vision, the very idea of using animal models to investigate human color vision presupposes an evolutionary commitment, and has been remarkably fruitful (Zeki 1993). In this book, following in Humphrey's footsteps, I make the case for the evolution of a mindreading capacity, basing it in part on the evidence of a genetic pathology that causes certain individuals to be born mindblind.

Evolutionary psychology looks at the brain (and thus the mind) as an organ that, via natural selection, has evolved specific mechanisms to solve particular adaptive problems. Darwin (1872) began this approach to psychology. As the examples in the previous paragraphs suggest, how to transmit information between individuals is likely to have been one of the adaptive problem that faced early hominids, and a language center in the brain is a solution to this problem. How to distinguish objects of identical shape and size (e.g., poisonous versus edible berries) is likely to have been another problem facing our evolutionary ancestors; a color-vision center in the brain is an efficient solution to this problem. Similarly, an attachment system solves the adaptive problem of ensuring that immature members of the species survive. Cosmides et al. (1992) use the metaphor of the brain as a Swiss Army knife to make this point. Each blade of the Swiss Army knife was, clearly, designed for a specific purpose: the corkscrew for pulling corks, the screwdriver for driving screws, the scissors for cutting thin materials, the saw for sawing thicker materials, and so on. It makes no sense to try to use the corkscrew to drive screws, or the screwdriver to pull corks, if the knife has a different "mechanism" for solving each of these problems. So it is with the brain, say Cosmides et al. We do not use our color-vision system to talk, or our language system to see color. We use specialized modules for the functions they evolved to solve.[2]

Evolutionary psychology, then, aims to account for the functioning of specific cognitive mechanisms and processes in humans. It also aims to account for the neurobiology, the adaptive value, the phylogenesis, and the ontogenesis of these mechanisms. Finally, it aims to describe any pathologies of these mechanisms. It is worth keeping our sights squarely on this goal. (Most psychological theorists would probably wish to be this broad-ranging, but it is often only those theories that are cast in the evolutionary framework that succeed in keeping their focus on all these separate levels.)

A final qualification about the use of the term "biological": It does not follow that only universals are biological. Individual differences can be biological, too. (Classical Darwinian theory depends on individual differences as something for natural selection to work on.) Similarly, just because something is biological, it does not follow that it is innate. And just because something is innate, it does not follow that it has a modular structure. Finally, even if something is innate and modular, whether it was caused by natural selection is still an open question. (For example, Pinker [1994] clearly disagrees with Chomsky's view that language is innate and modular but not necessarily the product of natural selection.) Claims that various behaviors have innate, biological, modular bases are separable and need independent justification.

In this book, I try to approach the goal of evolutionary psychology with regard to one specific adaptive problem: the rapid comprehension and prediction of another organism's behavior. I aim to sketch how mindreading may be the solution to this problem, and how mindreading has an innate, biological, modular basis. Cosmides et al. (1992, p. 8) define an adaptive problem as a "problem whose solution can affect reproduction, however distally,"[3] since the mechanism that shapes the evolution of a biological system is the system's contribution to the individual's success at reproducing (which enables transmission of the genes that built the biological system in the first place). This, of course, is the familiar concept of natural selection. It is evident, I think, that if another organism's next action is going to be to attack you, or to share its food with you, or to mate with you, you would do well to anticipate this quickly, since any of these actions could indeed "affect reproduction, however distally."

To summarize: One key aim of evolutionary psychology is to describe the evolution of neurocognitive mechanisms. To do this requires the theorist to have some notion of what biologists such as Bowlby (1969) call the *environment of evolutionary adaptedness* (EEA). This is the environment to which the mechanism was an adaptation. In the case of the evolutionary psychology of mindreading, what might have been the relevant EEA?

The Social Environment of Evolutionary Adaptedness

Clearly, the period of recent human history is not the relevant EEA. Cosmides et al. (1992, p. 5) remind us that nothing about our biology—not even our mindreading capacity—is likely to be an adaptation to modern human history. This period is only about 10,000 years old (if one dates it from the advent of agriculture, for example). In contrast, the period relevant to human evolution can be seen as spanning two phases: the Pleistocene epoch (roughly the last 2 million years, during which humans lived as hunter-gatherers), and the several hundred million years before that (during which they lived as foragers of one kind or another). These time spans are important to keep in mind because "they establish which set of environments and conditions defined the adaptive problems the mind was shaped to cope with: Pleistocene conditions, rather than modern conditions" (ibid., p. 5).[4]

That there was massive neurocognitive evolution during the Pleistocene epoch is beyond any doubt. The brain has increased threefold in size in the 3 million years since *Australopithecus afarensis* evolved, going from around 400 cubic centimeters to it current size of about 1350 cubic centimeters (figure 2.1).

The increase in brain size is likely to have had many causes, but one key factor upon which many theorists (Humphrey 1984; Byrne and Whiten 1988; Cosmides 1989; Brothers 1990) agree is the need for greater 'social intelligence'—shorthand for the ability to process information about the behavior of others and to react adaptively to their behavior. It is likely that there was a need for greater social intelligence because the vast majority of

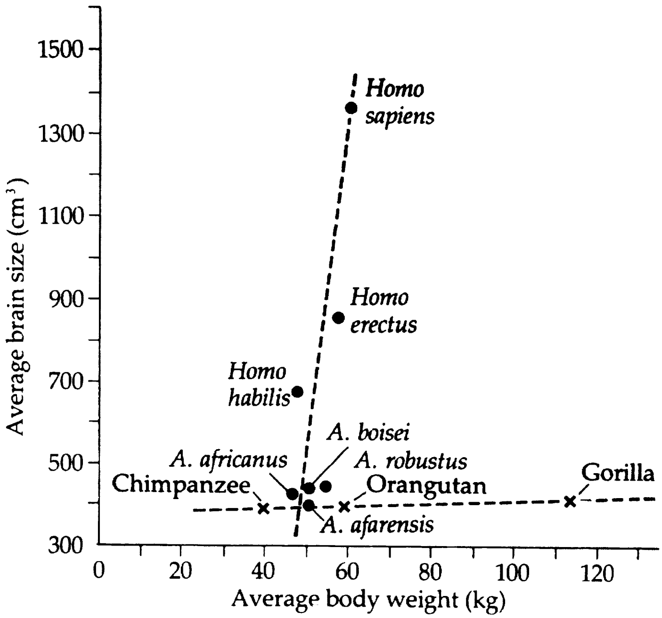

Figure 2.1
Changes in brain size over the last 3 million years. Reproduced from Lewin 1992.

non-human primate species are social animals, living in groups that range from as few as two individuals to as many as 200.

If you are living in a social group of two, the complexities of behavior that you need to make sense of are at least as demanding as our initial example ("John walked into the bedroom"). If you are living in a group of 200, making sense of the social behavior is staggeringly complex. One needs a powerful device—or set of devices—to make sense of actions, rapidly, in order to survive and prosper.

Though we cannot be sure what the EEA for social intelligence was, the striking variety of social organizations in existing non-human primates is a clue. Some primates (e.g., the gibbon) are monogamous. Some (e.g., the gorilla) live in what is called "unimale polygyny," where a single male has control over a group of females and their offspring. Others (e.g., the chimpanzee) live in "multimale polygyny," where several males cooperate to defend a group of widely distributed females and their offspring. Finally, some (e.g., the orangutan) live in what Lewin (1992) calls "exploded unimale polygyny," where a single male defends a group of females and their offspring but the females do not live as a group and instead are distributed over a wide area. If we assume that all these primates evolved from common ancestors, it is plausible that increasing social complexity was an adaptive problem facing them. It may have been as a result of changes in the brain that the complexity of social organization increased, or perhaps new brain mechanisms dedicated to social intelligence evolved to solve the problems posed by the increasing complexities of social behavior. Either way, we are led to look closely at the brain basis for social intelligence (Dunbar 1993).

But let us leave the brain for later and return to the question of social behavior and the EEA. Again, primates are instructive. As the various models suggest, primates live in social environments that are much more complex than the social environments of other animals that live in groups of comparable size, such as the antelope, the sheep, and the cow. (Have you ever noticed how rarely sheep interact?) The difference in complexity lies in the nature of the social interactions. As Lewin (1992, p. 46) puts it: "The [primate] group is . . . the center of intense social interaction that has little apparent direct bearing on the practicalities of life: in the human sphere we would call it socializing, the making and breaking of friendship and alliances."

The challenge for the primate was (and remains) to understand, predict and manipulate the behavior of others in the group. Byrne and Whiten (1988) depict this as the Machiavellian nature of social interaction: to interact in order to use others for various purposes. In primate groups it is this social intelligence that determines who wins higher status. Consider Lewin again on this point:

> When you observe other mammal species and see instances of conflict between two individuals, it is usually easy to predict which one will triumph: the larger one, or the one with the bigger canines or bigger antlers, or whatever is the appropriate weapon for combat. Not so in monkeys and apes. Individuals spend a lot of time establishing networks of "friendships," and observing the alliances of others. As a result, a physically inferior individual can triumph over a stronger individual, provided the challenge is timed so that friends are at hand to help the challenger while the victim's allies are absent. (Lewin 1992, p. 129)

The paleontologist Richard Leakey, writing with Lewin, reaches a similar conclusion:

> The world of higher primates—of monkeys, apes, and humans—is quintessentially a game of social chess, a keen intellectual challenge. The challenge is keener yet than the ancient board game itself, because the pieces not only unpredictably change identity—knights becoming bishops, pawns becoming castles, and so on—they occasionally switch colors to become the enemy. . . .
>
> What each individual seeks, of course, is reproductive success: producing as many healthy, socially adept offspring as possible. In birds of paradise, the greatest reproductive success (in males) goes to those with the most elaborate plumage and winning display. In red deer, the greatest reproductive success (again, in males), goes to those with the biggest, strongest bodies with which to overthrow rivals, sometimes literally. In higher primates, the greatest reproductive success (in both males and females) is shaped much more by social skills than by physical displays, either of strength or appearance. The complex interactions of the primate social nexus serve as an exquisite sorting system, in which the individuals with an edge in making alliances and monitoring the alliances of others may score significantly in reproductive success. (Leakey and Lewin 1992, pp. 191–293)

So, one view is that the evolution of primates is characterized by an increase in the complexity of social interaction, requiring (on the cognitive level) an increase in rapid and adaptive social intelligence and (on the biological level) an increase in different brain mechanisms to support this. The "Machiavellian hypothesis" of brain evolution is by no means proven, since alternative hypotheses are not ruled out by current evidence; it seems a strong contender, however.

Social Chess

The metaphor of "social chess" was Humphrey's. His idea was that intelligence evolved to enable organisms living in complex social groups to understand and take advantage of community living. In Humphrey's words (1984, p. 19), "the chief role of creative intellect is to hold society together." This proposal assumes a sharp distinction between social intelligence and other kinds of intelligence. Here is Humphrey on this:

> "Social intelligence" required, for a start, the development of certain abstract intellectual skills. If men were to negotiate the maze of social interaction it was essential they should become capable of a special sort of forward planning. They had to become calculating beings, capable of looking ahead to yet unrealized possibilities, of plotting, counter-plotting and pitting their wits against group companions no less subtle than themselves. Never before, in their dealings with the non-social world, the world of sticks and stones, not even in their dealings with the world of living predators and prey, had human beings needed their powers of abstract reasoning which they now needed in their dealings with each other. But now their very survival within the social group depended upon it. . . . The life of social animals is highly problematical. In a complex society, such as those we know exist among higher primates, there are benefits to be gained for each individual member both from preserving the overall structure of the group and at

the same time from exploiting and out-maneuvering others within it. Thus social primates are required by the very nature of the system they create and maintain to be calculating beings; they must be able to calculate the consequences of their own behavior, to calculate the likely behavior of others, to calculate the balance of advantage and loss—and all this in a context where the evidence on which their calculations are based is ephemeral, ambiguous, and likely to change, not least as a consequence of their own actions. In such a situation, "social skill" goes hand in hand with intellect, and here at last the intellectual faculties required are of the highest order. The game of social plot and counter-plot cannot be played merely on the basis of accumulated knowledge, any more than can a game of chess.

Like chess, social interaction is typically a transaction between social partners. One animal may, for instance, wish by his own behavior to change the behavior of another; but since the social animal is himself reactive and intelligent the interaction soon becomes a two-way argument where each "player" must be ready to change his tactics—and maybe his goals—as the game proceeds. Thus, over and above the cognitive skills which are required merely to perceive the current state of play (and they may be considerable), the social gamesman, like the chess-player, must be capable of a special sort of forward planning. Given that each move in the game may call forth several alternative responses from the other player this forward planning will take the form of a decision tree, having its root in the current situation and branches corresponding to the moves considered in looking ahead at different possibilities. It asks for a level of intelligence which is, I submit, unparalleled in any other sphere of living. There may be, of course, strong and weak players—yet, as master or novice, we and most other members of complex primate societies have been in this game since we were babies. (Humphrey 1984, pp. 4, 20–21)

The metaphor of social chess works well on several levels. Like chess, social interaction requires one to have a strategy in order to get gradually closer to one's goal, to keep track of the changing positions of various participants and how these affect one's own position, to figure out what might happen next and anticipate it by reacting accordingly, and to outwit one's opponent. It also reminds us that social interaction can be as challenging in problem-solving terms as is chess: just as it is not self-evident exactly where to place one's pieces so that your opponent's key piece cannot move, it is not self-evident how to predict a person's actions, or how to get what you want from another person, or how influence a number of individuals in a group.

The metaphor of social chess risks distorting the concept of social interaction in some ways, too. First, not all social interaction is competitive in the way that chess is. (Even cooperative social interaction requires considerable mindreading.) Second, for many of us, actually playing chess can be rather laborious—a solution to whatever fix one is in does not just pop into one's head. Indeed, in most chess tournaments there is a time limit, since the computing of a solution can be frustratingly slow. In contrast, for most of us, judgements about action in a social situation and about how to get on in a social group do not typically involve laborious logical reasoning. The way we play social chess seems far more intuitive. We just know what to do, and we can easily surmise the reasons that might lie behind someone's actions.

Expert chess players may feel that they intuitively know what the next best move is, and their skill at chess may be an excellent metaphor for how we routinely make judgements during social interaction. Like the chess expert, we are social experts. Our social reasoning process has become automatic and effortless—possibly as a result of years of daily practice, possibly also because, right from the beginning of life, the human brain is programmed to automatically and effortlessly interpret social behavior in this way, as a result of millions of years of evolution. Perhaps we never go through a stage of finding social interaction an effort to decode. Rather, we are born understanding social chess, or at least we have many of the basic principles that we will need in order to make sense of and take part in the game. We have some key neural mechanisms that allow us to "see" the solution to a social situation intuitively.

Chapter 3

Mindreading: Nature's Choice

> In evolutionary terms it must have been a breakthrough. . . . Imagine the biological benefits to the first of our ancestors who developed the ability to make realistic guesses about the inner life of his rivals; to be able to picture what another was thinking about and planning to do next; to be able to read the minds of others by reading his own. (Humphrey 1984)

Before I describe the specific mechanisms that I think evolved to enable us to mindread, I have to do one last bit of stage setting. I want to persuade you that mindreading is simply the best way to make sense of the actions of others. If I can persuade you of this, then it will become clearer why evolution may have favored this solution.[1] It will, I think, be easy to convince you, since I think that the alternatives to mindreading do not even come a close second in the competition. I will rely mainly on an argument of Dennett's (1978a) to make this case.

Here, then, is why mindreading is a wonderful thing for us to have: Attributing mental states to a complex system (such as a human being) is by far the easiest way of understanding it. By understanding, Dennett means coming up with an explanation of the complex system's behavior and predicting what it will do next. Dennett calls this ability "adopting the Intentional Stance." The term "Intentional Stance" refers to our ability to attribute the full set of intentional states (beliefs, desires, thoughts, intentions, hopes, memories, fears, promises, etc.), not just to the specific mental state of intention.

What Are the Alternatives to Mindreading?

Dennett argued that the two alternatives to adopting the Intentional Stance are attempting to understand systems in terms of their physical makeup (the Physical Stance) and attempting to understand them in terms of their functional design (the Design Stance).[2] We adopt the Physical Stance to understand systems whose physical makeup we know about. For example, we all know a bit about human anatomy. We know enough to be able to reason that the skin bleeds when it is cut because blood vessels have been severed. In this instance, our understanding of the body takes the form of adopting the Physical Stance—using a "folk biology." However, attempting to understand the behavior of a person or of any kind of animal by adopting the Physical Stance would be a non-starter, in view of the state of our knowledge: we would need to know about the millions of different physiological (brain) states that give rise to different behaviors in order to understand animal or human behavior in purely physical terms.

Mindreading (or adopting the Intentional Stance) is therefore an infinitely simpler and more powerful solution than adopting the Physical Stance. Thus, in response to the first example in chapter 1, the answer

> John went into the bedroom because he **wanted** to find his jacket and **thought** it was in there

may be just as accurate as

> John went into the bedroom because six **brain states** (A, D, F, H, J, and Q) were activated in a particular sequence (D, J, Q, A, F, and H);

however, the former is far simpler to compute. The latter account assumes that John's current behavior is due to just six brain states, active in a particular sequence. In reality, of course, it is likely to be a vastly larger number than this. This Physical Stance account also assumes that these brain states are known, which at present they are not. Finally, it assumes that these brain states are knowable, whereas in reality neither ordinary folk nor scientists have a "brainoscope" with which to "see" what brain states are active in another person at any given time and then to predict that person's next action.

So much for the first of the alternatives to the Intentional Stance: the Physical Stance is good for some things, but it is not up to the job of predicting the behavior of complex systems. But what of the second alternative? According to Dennett, we adopt the Design Stance when we are ignorant about the physical makeup of a system (the Physical Stance is therefore not available to us) but are trying to understand the system in terms of the functions of its observable parts. For example, I need not know anything about silicon chips or other details of my computer's physical makeup to predict its behavior. Instead, I can refer to some of its design features: the Delete key (whose function is to rub out what I have just typed), the Escape key (whose function is to clear the screen), and so forth. Not only is the Design Stance explanatory (to a degree); it also allows us to predict what the system will do next if this button or that switch is pushed.

The Design Stance works well when we wish to explain a system composed of clearly observable and operational parts, such as an alarm clock, a television, or a thermostat. Note, however, that mindreading would work just as well. Indeed, many people reason about their computers in very mentalistic ways, saying things like

> My computer is displaying this command because it **thinks** I have finished.

Dennett reminds us that this is how we talk about thermostats too:

> It **wants** to keep the room at a constant temperature, and **thinks** that the room is getting too warm.

However, the Design Stance seems just as useful in these cases. For example, we might say

> My computer isn't working because it is not plugged in.

But adopting the Design Stance toward understanding the behavior of people or other animals would not get one very far, since people and animals have very few external, operational parts for which one could work out a functional or design description. It might work well enough for making sense of people's reflexes (e.g., an eye blinks when you blow on its surface or poke it with your finger, because it has a pressure detector on its surface whose function is to close the eye). The Design Stance is also pretty useful for studying unobservable processes; however, it is of little value in making sense of and predicting moment-by-moment changes in a person's behavior.

For Dennett, we use mindreading because it works. It is a pragmatic argument. Dennett is not committed either way on the question of whether there really are such things as mental states inside the heads of organisms. We ascribe these simply because doing so allows us to treat other organisms as rational agents. For Fodor (1983), it is not just a pragmatic issue. We mindread, according to him, because there really are mental states inside our own and other organisms' heads. Inferring these states gives us a powerful lever in making sense of and predicting behavior. (In contrast, the Churchlands [1981] acknowledge that we employ mindreading but argue that the idea that mental states actually exist is patently false.)

Let us just focus on Fodor's view that the solution evolution came up with to enable us to understand and predict our own and other people's behavior—or the behavior of any complex system—is the Intentional Stance, or what we have so far been calling mindreading. As Fodor points out, it is a simple-to-use and powerful theory, which is exactly what we need when we are in the thick of a social situation. It also leaves the alternatives miles behind. Evolution was not going to wait around for human scientists to invent a brainoscope before primates (early hominids included) could understand and participate in complex social interaction. If it had done so, the hominid line would have died out long ago. Instead, I argue, it gave us a far simpler device, something akin to a mindoscope, and it gave it to us as part of our neural anatomy, which allows us to mindread other creatures.[3]

Just to make this more real, imagine that a salesman is standing at your front door, waiting for you to sign a piece of paper. You need to reason quickly about his behavior and what he is likely to do next. Making inferences about his desires, intentions, thoughts, and motives allows you to do this. Now imagine that you are an early hominid, and that another early hominid offers to groom you and your mate. You need to reason quickly about whether you should let him approach. Again, making inferences about whether his motives are purely altruistic or whether he might be deceitful is a reasoning strategy that you can apply in time to react to a social threat. The modern human, at least, seems to be extremely quick at mindreading, and appears to engage in it automatically. My guess is that hominids have been doing this for a long time.

Mindreading also goes under the name "folk psychology"—and that may be a better term for it,[4] since it reminds us that it is simply our everyday way of understanding people. As Dennett (1987, p. 48) points out,

> We use folk psychology all the time, to explain and predict each other's behavior; we attribute beliefs and desires to each other with confidence—and quite unselfconsciously—and spend a substantial portion of our waking lives formulating the world—not excluding ourselves—in these terms. . . . Every time we venture out on the highway, for example, we stake our lives on the reliability of our general expectations about the perceptual beliefs, normal desires and decision proclivities of the other motorists. We find . . . that it is a theory of great generative power and efficiency. For instance, watching a film with a highly original and unstereotyped plot, we see the hero smile at the villain and we all swiftly and effortlessly arrive at the same complex theoretical diagnosis: "Aha!" we conclude (but perhaps not consciously), "He wants her to think he doesn't know she intends to defraud her brother!"

A final alternative to mindreading is what might be called "adopting the Contingency Stance." This entails learning or innately recognizing the behavioral contingencies between another organism's behavior and their effects. For example, seeing a cat arch its back might make you anticipate that it is about to pounce. Seeing a gorilla flare its nostrils, open its eyes and mouth wider, and beat its chest might make you anticipate that it is about to attack. The Contingency Stance is probably a subspecies of the Design Stance. In essence, to adopt this stance is to characterize the organism as a behaviorist—the most obvious alternative to a mindreader. It is highly likely that most nonhuman organisms adopt this stance in their interactions with other animals, and that we humans often make use of it too. Seeing someone yawn in mid-conversation, you might anticipate that the conversation will be ending very soon; seeing someone raise his fist might act as a powerful cue for you to flinch and take self-protective measures. Darwin's studies suggested that a long line of organisms display, recognize, and react to bodily expressions. However, picking up behavioral cues is obviously useful only if such cues are available. What mindreading allows one to do is predict behavior even in situations where there are no behavioral cues. For example, not having heard from a friend for several months, you might think that she thinks you have offended her in some way; you might then decide to give her a ring to check how things are. The friend's lack of behavior can hardly be considered a specific "cue," but a mindreader can work with only minimal behavioral cues (and even with none whatsoever—one can attribute mental states to an invisible entity, like God).

Mindreading and Communication

So far, in attempting to persuade you that mindreading is a really good thing to have, I have focused on its role in making sense of behavior. However, a second reason why mindreading is useful, and thus why it may have evolved, is the way in which it allows us to make sense of communication. Let me add to my argument about the virtues of mindreading by saying just a little about its role in communication.

A range of theorists—Grice (1967), Sperber and Wilson (1986), Austin (1962)—have argued that when we hear someone say something (or when we read a sentence in a novel), aside from decoding the referent of each word (computing its semantics and syntax), the key thing we do as we search for the meaning of the words is imagine what the speaker's communicative intention might be.[5] That is, we ask ourselves "What does he mean?" Here the word "mean" essentially boils down to "intend me to understand." Put another way, the key question that guides our comprehension process is "Just what is he driving at?" The notion is that not only do we pay attention to the actual words a speaker uses; we also focus on what we think was the gist of what he or she wanted to say or wanted us to understand. Sperber and Wilson (1986) call this a search for "relevance"—the listener assumes that the meaning of an utterance will be relevant to the speaker's current intentions. Thus, when the cop shouts "Drop it!" a robber is not left in a state of acute doubt over the ambiguity of the term "it." Rather, the robber makes a rapid assumption that the cop meant (i.e., intended the robber to understand) that the word "it" should refer to the gun in the robber's hand. And at an even more implicit level, the robber rapidly assumes that the cop intended the robber to recognize his intention to use the word in this way. In decoding figurative speech (such as irony, sarcasm, metaphor, or humor), mindreading is even more essential.

The above analysis of language makes clear that in decoding speech we go way beyond the words we hear or read, to hypothesize about the speaker's mental states. This analysis applies not only to speech but also to non-verbal communication. Thus, when I gesture toward a doorway with an outstretched arm and with an open palm, you immediately assume that I mean (i.e., intend you to understand) that you should go through the door.

A more complex example (from Sperber and Wilson 1986) makes the point of how we search the context in order to infer a speaker's communicative intention even more clearly:

Flag-seller: Would you like to buy a flag for the Royal National Lifeboat Institution?

Passerby: No thanks, I always spend my holidays with my sister in Birmingham.

Sperber and Wilson's analysis of this brief exchange runs as follows:

> To see the relevance of the passerby's response, the hearer must be able to supply something like the premises [below] . . . , and derive something like the contextual implication [below]:
>
> (a) Birmingham is inland.
> (b) The Royal National Lifeboat is a charity.
> (c) Buying a flag is one way of subscribing to a charity.
> (d) Someone who spends his holidays inland has no need of the services of the Royal National Lifeboat Institution.
> (e) Someone who has no need of the services of a charity cannot be expected to subscribe to that charity.
>
> [Therefore,] the passerby cannot be expected to subscribe to the Royal National Lifeboat Institution. (ibid., pp. 121–122)

In some ways, what the passerby said was rather unconnected. Nevertheless, there is still a way we can work out why he might have said it and what he might have wanted to mean: by representing all of the above thought steps that could have been in his mind. Our mindreading fills in the gaps in communication and holds the dialogue together.

Let us take just one other example of this, this one from Pinker (1994, p. 227):

Woman: I'm leaving you.

Man: Who is he?

In order to have produced this phrase, the man must have thought that the woman was leaving him for another man. When we make this attribution to the man, the dialogue hangs together perfectly. If we did not make it, the dialogue would seem disconnected, almost a random string of words. As mindreaders, we perceive the man's sentence as far from random.

Presumably, a person with mindblindness would struggle in vain to find any relevance in this exchange.

The other way in which mindreading is held to play an essential part in successful communication is in the speaker's monitoring the informational needs of the listener—that is, in the speaker's judging what the listener may already know or be ignorant about, and what information he or she should supply so that the listener will be able to understand the message. An utterance like "Shall we?" will not, in most contexts, be enough to get the listener to know what on earth you were intending to say. (A dance hall might be an exception in the case of this particular phrase.) Rather, the speaker computes that, if the listener is to have even a glimmer of an idea of what he is trying to say, he must provide more information—for example, "Shall we go to Devon this weekend?"

Furthermore, for communication to succeed, the speaker must monitor whether the meaning of an utterance has been received and understood as he or she intended it to be, or whether rephrasing is required to resolve ambiguity. Dialogue understood in this way becomes much more than the production of speech: it is revealed as intrinsically linked to the use of mindreading skills. The alternatives to mindreading (such as adopting the Physical or the Design Stance, or using what Sperber and Wilson (1986) call a "code" approach to language) turn out to be inadequate means of making sense of human communication and action. And in terms of the evolutionary "fitness" value of mindreading in communication, it is clear that an ability to go beyond the words heard to the speaker's intention would allow the listener to judge whether the speaker was being deceitful or genuine.

A final sense in which language and mindreading are intimately related rests on the idea that language functions principally as a "printout" of the contents of the mind. We speak to one another to share our ideas, thoughts, and experiences. Whether mindreading came first, and language evolved to facilitate this, or the other way around, remains unclear.

To summarize: Mindreading is good for a number of important things, including social understanding, behavioral prediction, social interaction, and communication. The lack of competitive alternatives to mindreading that could produce equal or better success in these domains makes it clearer why natural selection might have latched onto mindreading as an adaptive solution to the problem of predicting behavior and sharing information. I mean, what other real choice did Nature have?