1 Mindreading in Animals: Its Importance and History

The question of whether animals read minds may at first sound somewhat comical, evoking images of telepathic pets with paws pressed meditatively to their temples, perhaps. However, the term 'mindreading' used in this book and in cognitive science and philosophy generally departs considerably from its more familiar meaning of direct mind-to-mind communication.[1] Here the term is used to refer to the ability to attribute mental states, such as beliefs, intentions, and perceptual experiences, to others by the decidedly mundane and indirect means of observing their behaviors within environmental contexts.

On this use of the term, ordinary human beings (and not just telepathists) are mindreaders par excellence. We quite readily predict and make sense of other humans' and animals' behaviors by trying to understand what they might be thinking, or seeing, or intending to do. Although we may actually mindread less than we typically think and say we do, as some researchers have argued (Hutto 2008; Bermúdez 2009), there are many everyday cases, as well as countless scientific studies, that quite clearly show that humans have an almost reflexive tendency, beginning early in childhood, to predict, understand, explain, and manipulate the behaviors of others in terms of what we think is going on in their minds (Wellman 1990; Carruthers & Smith 1996; Nichols & Stich 2003; Goldman 2006; Gopnik 2009). Regarding mindreading in humans, there is simply no serious issue.

By contrast, there is a deep and important question in cognitive science and philosophy concerning whether humans are the only living animal capable of mindreading. For over thirty years now, there has been a sustained and heated debate over the question of whether any species of nonhuman animal is capable of attributing mental states. The field is presently divided between those who claim that there is empirical evidence to support the hypothesis that some animals are mindreaders and those who deny this.

In my opinion, the animal mindreading debate has simply stalled, with both sides digging in their heels. This is not to say that empirical research into animal mindreading has stalled—quite the opposite, in fact. In the past eight years or so, there has been an unprecedented flurry of empirical research into mindreading capabilities in a variety of species of animal. Rather, the debate has stalled in that both sides are content with a handful of arguments that, I believe, fail to support the rather strong claims that they are used to make. Contrary to what either side holds, the question of whether animals are mindreaders is in no way a settled matter.[2] I do believe, however, that there is a way to move the debate closer to a settled answer. I shall make the case that what is fundamentally needed to advance the debate is a reconceptualization of what mindreading (specifically, cognitive state attribution) in animals should be understood to encompass and (most important) a solution to an epistemic and methodological problem (dubbed the 'logical problem') that has dogged the field since its inception.

It should be noted at the start that in this debate the mentality of animals as such is not in dispute. It is recognized on all sides, and for good scientific and commonsensible reasons, that many animals have minds, that they perceive, feel, have emotions, perform intentional acts, have beliefs, remember, and (in some instances) engage in various forms of cognition and reasoning (see Griffin 1992; Shettleworth 1998). Thus, it is acknowledged by both sides that animals live in a world with other minded creatures—a world of perceiving, feeling, intending, thinking conspecifics, owners, predators, and prey. And it is equally recognized that animals are adept at predicting the behavior of these other creatures. This is all common ground. The questions that define the debate, rather, are (i) whether animals ever make such predictions by understanding what is (or what they take to be) going on in another creature's mind or simply by being sensitive to the behavioral and environmental cues that indicate (or can be used to infer) what the other creature is likely to do without interpreting them as evidence of the creature's state of mind and (ii) whether behavioral science, with its empirical methods of investigation and testing, has the resources to determine which of these hypotheses is correct.

This book is about this debate, why it is important, and how it can be moved forward. In section 1.1 below, I state some of the reasons the question of animal mindreading is important to cognitive science and philosophy. In section 1.2, I give a concise summary of the history of the debate, its current state of stalemate, and what, in general terms, needs to be done to advance it.

1.1   Why the Question of Animal Mindreading Matters

The question matters for three general types of reasons: those related to our interest in the minds and moral status of animals; those related to our interests in certain philosophical theories about mindreading and the nature of the mind; and those related to various scientific interests in studying mindreading in humans. Below is a representative set of these different concerns.

Relevance to the scientific study of animal minds   Cognition is pervasive in the animal world, and scientists and philosophers have had an abiding interest in studying it, not only for the light it can shed on human cognition, but for what it can tell us about the minds of animals (Walker 1983; Gallistel 1990; Griffin 1992; Shettleworth 1998; Bekoff et al. 2002; Lurz 2009b). Although scientific investigation into animal mentality lay largely dormant during much of the heyday of behaviorism, it has seen a healthy rebirth in the past thirty years or more. Under the auspice of the cognitive revolution in psychology and philosophy, researchers from various fields are now vigorously pursuing questions about animal minds, and one of the more active research programs in the field is the question of mindreading (and social cognition more generally) in animals. Researchers from cognitive and comparative psychology, cognitive ethology, and cognitive neuroscience have been investigating various social cognitive capacities in different animal species (e.g., apes, monkeys, dolphins, dogs, goats, pigs, elephants, and birds) in part for what it can tell us about how these animals understand their social worlds, how this social knowledge may have evolved within a given species, how it compares across different species, and how it may develop within the individual members of the species.

In addition, the question of mindreading in nonhuman primates (hereafter primates) is of particular importance to an influential hypothesis about the evolution of primate intelligence called the 'Machiavellian intelligence hypothesis' or 'social brain hypothesis' (see Jolly 1966; Humphrey 1976; Byrne & Whiten 1988; Whiten & Byrne 1997; Dunbar 1998, 2007). According to the hypothesis, it was a set of demanding social problems resulting from living in large, complex groups that drove the evolution of primates' large neocortex (the seat of higher cognition) relative to their body size and their correspondingly unique intellectual abilities, such as tool use, coalition and alliance formations, mirror self-recognition, and the capacity for transitive inferences, among others. To survive and thrive in their complex social world, according to the hypothesis, primates evolved

certain mindreading abilities needed for solving complex social problems, such as anticipating and manipulating conspecifics' behaviors. A number of researchers have now extended the Machiavellian intelligence hypothesis to explain the evolution of similar forms of higher intelligence and brain size found in certain nonprimate animals, such as corvids and dolphins, that also live in large, complex social groups. To date, the Machiavellian intelligence hypothesis (or social brain hypothesis) appears to fit the neurological data—that relative brain size correlates with group size and social complexity—better than any of the alternative ecological hypotheses of the evolution of intelligence (Dunbar 2007). Of course, whether the hypothesis is correct regarding the evolution of mindreading in these social species of animals depends critically upon whether these animals really do have the capacity to attribute mental states.

Relevance to the moral status and welfare of animals   Philosophers and ethicists make an important distinction between the biological category of human being, *Homo sapiens*, and the moral/psychological category of persons. *Homo sapiens* are defined by their unique genetic make-up, ability to interbreed, morphology, and evolutionary history, whereas *persons* are defined by a unique set of moral and psychological attributes and capacities, such as the capacity for moral autonomy (roughly, the ability to understand right and wrong and act for moral reasons), self-awareness, and mindreading. Although all (or most) *Homo sapiens* are persons, it does not follow that only *Homo sapiens* are persons, or even legal persons (i.e., entities that are recognized by law as persons).

It has been recently argued by some philosophers and scientists that since the great apes share enough of their DNA, evolutionary history, and psychological capacities with human beings, they deserve to be recognized as persons or (at the very least) as legal persons. Often cited on this list of relevant psychological similarities is the ability of great apes to attribute mental states (or to 'empathize'). And this is no accident. For there is a well-established line of argument in philosophy and cognitive science that holds that of the attributes that define personhood, mindreading is the most central (Dennett 1978b). It is mindreading, the argument runs, upon which empathy and (subsequently) moral autonomy rest, as well as self-awareness, and it is mindreading that should be used as one of the logical grounds for distinguishing animals that should be recognized as (legal or natural) persons from those that should not (see Mitchell 1993). It is important to note that on the basis of such arguments, Spain's parliament recently approved a resolution for the legal recognition and protection of

the rights of life, liberty, and freedom from torture for the great apes (Abend 2008). Other countries in Europe seem poised to follow suit.

Independent of the issue of the status of animals as persons, the question of animal mindreading has recently been argued to be relevant to issues of animal welfare (Tschudin 2001). The argument is that if some animals are mindreaders, then they likely have a natural tendency to predict and understand conspecifics' behaviors in terms of the mental states they attribute, and thus they may well be harmed by being placed in conditions that either prevent this natural tendency from being exercised or exercised in ways that induce stress. For example, it has been recently argued by Held et al. (2001) that domestic pigs, which have shown some tentative signs of being capable of understanding the visual perspective of conspecifics, may be adversely affected by the common husbandry practice of mixing these animals in with unfamiliar conspecifics (which would be difficult for them to mindread) or from their being isolated from familiar conspecifics (which would prevent them from exercising their natural tendency to mindread), as well as their being subject to witnessing familiar conspecifics suffering (which may cause them to empathize with the other's pain). It is argued that an improved understanding of mindreading abilities in animals may lead to improvements to their housing conditions and related husbandry practices.

Relevance to philosophical theories on mindreading and the nature of the mind   There are three influential theories in philosophy on the importance of mindreading in shaping a distinctively human mode of cognition. According to the first of these theories, which is traceable through Wittgenstein (1953) and Davidson (1980, 2001), mindreading of a certain type is required for the possession of genuine thought (cognitive states that can be literally true or false, correct or incorrect). The general idea behind the theory is that only through the ascription of thoughts to others can a creature come to grasp a notion of objectivity—the way things really are independently of its own mind—that is required to make sense of how its own cognitive states can be true or false. Defenders of this theory typically go on to argue that mindreading of this sort is only possible for linguistic creatures, such as normal adult human beings, and conclude that only humans are capable of genuine thought and reason. We shall return to this argument and examine it in more detail in chapter 4. For now, it is important to see that the question of animal mindreading is relevant to its plausibility.

A related line of thought, inspired by Peter Strawson (1959), is that mindreading of a certain type is required for possessing thoughts about

particular objects and their characterizing properties. According to this theory, nothing short of ascribing thoughts (e.g., a belief) and their epistemic grounds (e.g., the perceptual states that justify the belief) to the very same individual would force a creature to think in terms of re-identifiable particular objects or individual substances (for a related idea, see Welker 1988; Proust 2009). Deprived of this capacity, the argument runs, creatures can aspire only to what Strawson describes as a feature-placing mode of cognition, a way of thinking about the world in terms of repeatable and projectable kinds of stuff (e.g., rain, snow, gold) but not in terms of recurring individuals (e.g., individual conspecifics or individual objects). The defenders of this theory typically assume that since animals lack the ability to mindread, their manner of cognition is relegated to a feature-placing mode and is thus importantly different from the uniquely human mode of cognition involving thoughts about re-identifiable individuals and their characterizing properties (see Burge 2009, 2010 for a recent critical account of this theory). Again, the question of animal mindreading is relevant to whether this philosophical theory about a distinctively human mode of thought is correct.

According to the third theory (which has defenders in both philosophy and science), mindreading is a necessary condition for self-awareness and metacognition. The capacity to know or represent the contents of one's own mind, according to the theory, is dependent upon one's capacity to know or represent the contents of other minds. Self-awareness and metacognition are understood to be the result of turning one's own mindreading abilities onto oneself (Carruthers 2009). It is sometimes argued that since most (if not all) animals are incapable of mindreading, they are thereby incapable of self-awareness and metacognition. One influential defender of this view (Carruthers 2000) has gone on to argue that since a type of self-awareness (viz., dispositional higher-order thought) is necessary for phenomenal consciousness (the 'what-it-is-like' or sentient aspect of experience), animals lacking in any form of mindreading are thereby lacking in phenomenal consciousness. Obviously, whether some animals are mindreaders is quite relevant, according to this theory, to whether they are capable of self-awareness and phenomenal consciousness.

Relevance to human mindreading   The question of animal mindreading is of particular importance to a branch of evolutionary psychology that seeks to trace the origins of mindreading in human beings. Researchers in this field are interested in discovering the place in our evolutionary past where mindreading first emerged, the selection pressures that led to its

emergence, and how it has evolved from its inception. Various methods of research are used in this field, but a principal method is that of studying the higher primates (e.g., the great apes and monkeys), our closest living evolutionary relatives—for if any of the higher primates demonstrate mindreading capacities, then this would suggest that mindreading in human beings evolved within the primate line due to distinct selection pressures found there (Santos et al. 2007). Conversely, if primates fail to demonstrate any ability to mindread, then this would suggest that mindreading in human beings evolved within the hominid (human) line due to distinct selection pressure found there (Penn et al. 2008).

The scientific pursuit of the question of animal mindreading may also prove useful to scientists studying the mindreading capacities of nonverbal and preverbal humans, as well as those studying the neural basis of mindreading in humans generally. Scientists studying mindreading in aphasic adults and very young infants, for example, are in need of discriminating nonverbal mindreading tests, and, of course, scientists studying animal mindreading are in the business of designing such tests. It is quite possible, therefore, that advances made in the methods used to study mindreading in animals will prove useful for designing more effective or sensitive tests of mindreading in nonverbal and preverbal humans.

Scientific investigation into mindreading in primates may also prove valuable to neuroscientists investigating the neural architecture of mindreading in humans, as well as those interested in developing neuropharmacological treatments for mindreading disorders in humans, such as autism. Both types of investigations are in need of animal subjects to run more invasive and controlled studies that would be otherwise impractical or unethical to run on humans. Of course, the usefulness of such animal models will depend upon whether science is able to demonstrate that these animals are truly mindreaders.[3]

1.2   A Brief History of the Animal Mindreading Debate

Although the question of animal mindreading was a topic of interest in philosophy and science prior to the late 1970s, its official birth date is 1978, the year that David Premack and Guy Woodruff published their landmark paper, "Does the chimpanzee have a theory of mind?" In that paper, Premack and Woodruff described a series of innovative experiments with their chimpanzee, Sarah, the results of which, they argued, indicated that she was capable of attributing mental states to her trainers. In one

of these experiments, Sarah was shown a variety of videotapes of a trainer confronting different types of problems (e.g., reaching for a banana suspended from the ceiling or shivering by an unlit heater). The videos were then paused, and Sarah was given the opportunity to select from a pair of still photos. Both photos depicted the trainer engaged in an action, but only one of the photos depicted an action that would have solved the trainer's problem presented in the video. For example, after the suspended banana video, Sarah was presented with one photo showing the trainer stepping onto a chair below the area where the banana was suspended and another photo showing the trainer prone on the floor and reaching to the side with a stick. Sarah consistently chose the photo depicting the solution to the trainer's problem in the video in nearly every test trial. Her stellar performance led Premack and Woodruff to conclude that Sarah understood the trainer's behavior in the videos in terms of the trainer having a particular intention or goal (e.g., to reach the bananas or to be warmed by the heater) and, thus, selected those photos that depicted those types of actions that would lead to the satisfaction of the attributed intention or goal.

In the same paper, Premack and Woodruff described an innovative experimental approach for testing mindreading capacities in animals, subsequently called the knower–guesser protocol, which soon became a standard procedure in the field. In the knower–guesser protocol, a chimpanzee is required to discriminate between two trainers in order to receive or locate food hidden inside one of two containers. The chimpanzee, however, is prevented from observing which container has been baited but is allowed to observe that one of the trainers (the knower) has witnessed the baiting process while the other (the guesser) has not.[4] If the chimpanzee is a mindreader, Premack and Woodruff reasoned, then she ought to favor the knowledgeable trainer over the ignorant one in some way, either by choosing the former to indicate a particular container or by choosing the container indicated by that trainer. When Premack (1988) eventually tested some chimpanzees with the protocol, he discovered that the majority of them (3 out of 4) did in fact show a decided preference, from the very first trial, for choosing the knower over the guesser in selecting the baited container.[5] On the basis of such results, Premack (2007) has come to hold that chimpanzees are capable of attributing simple perceptual states and intentions to others, but not beliefs or knowledge, which he considers too complex for nonhuman animals to comprehend. This has become a rather common view among animal mindreading researchers and will be examined in more detail in chapter 4.

As important as Premack and Woodruff's paper was for initiating and structuring the direction of empirical research into animal mindreading, the critical commentary that accompanied their paper was equally important. A notable series of objections were leveled against Premack and Woodruff's interpretation of Sarah's performance in the video tasks. On closer inspection of the videos and photos, it was pointed out by some researchers (Savage-Rumbaugh et al. 1978; Heyes 1998) that Sarah could have chosen the correct photos simply on the grounds of some formerly learned associations (e.g., choosing a photo with a lit roll of paper because she associated the orange flame with the heater) or by matching items found in the last frame of the videos with the correct photos (e.g., choosing the photo with the horizontal stick since the stick was present in the final frame of the video). But by far some of the most significant commentary came from three philosophers: Gil Harman (1978), Daniel Dennett (1978a), and Jonathan Bennett (1978). All three independently pointed to a critical, underlying problem of empirically testing for mindreading in animals: the problem of experimentally distinguishing genuine mental state attribution in animals from the attribution of the observable cues or facts that serve as the grounds or evidence of such mental states in others. As we shall see, this problem, subsequently dubbed the 'logical problem' (Hurley & Nudds 2006; Lurz 2009a), has come to be held by a number of researchers as the main obstacle to answering the question of mindreading in animals.

The significance of their identifying this problem was matched by these philosophers' innovative proposals to solve it. Each recommended a radically new type of experimental paradigm—the false-belief test. The test was designed to determine whether an animal would anticipate another subject's behavior not simply on the basis of the objective, observable cues or facts regarding the other's behavior and environmental situation, but on the basis of what the other subject erroneously believed to be the case about its behavior/environmental situation. Harman, for example, suggested allowing one chimpanzee (A) to observe while another chimpanzee (B) watched an experimenter place food inside one of two containers. While chimpanzee B was momentarily distracted from the containers (e.g., it had its back to the containers), chimpanzee A would observe the experimenter remove the food from its original hiding place (i.e., the container in which chimpanzee B saw the experimenter place it) and place it in the other container. Harman (1978) reasoned that if chimpanzee A expected chimpanzee B (once it had turned around to face the containers) to look into the container that was originally baited (but

was no longer), then "that would seem to show that it has a conception of mere belief" (p. 577).

Although the false-belief test was ignored by animal researchers for many years after Premack and Woodruff's paper, it quickly became the litmus test for mindreading in developmental psychology. Wimmer and Perner (1983) were the first to use a version of the test on children. In their experiment, children from three to five years of age were shown a puppet show in which the main character (Max) had a mistaken belief about the location of his candy. It was discovered that five-year-olds (and some older four-year-olds) were capable of predicting Max's behavior by attributing a mistaken belief to him but that younger children were not. The younger children typically predicted Max's behavior in terms of where the candy actually was, not where Max mistakenly thought it was.[6] Wimmer and Perner's findings were quickly replicated and shown to be robust, and so, not surprisingly, there was a growing sense among some animal researchers that if three- and four-year-old children were incapable of attributing beliefs, animals were unlikely to do any better (see Premack 1988). This suspicion appeared to be borne out later by the negative findings from the first false-belief study on chimpanzees by Call and Tomasello (1999). Such results have, in turn, refueled the view, first held by Premack, that chimpanzees are capable of mindreading for simple states of perception and goal-directed/intentional actions but not for beliefs. However, as we shall see again and again, the road to discovering animal mindreading is anything but a straight line. O'Connell and Dunbar (2003), for example, have recently received positive results from a single chimpanzee on a false-belief test, and Tschudin (2001, 2006) has received even stronger results from three dolphins using a similar test. These studies will be discussed further in chapter 4.

At around the time of Premack and Woodruff's paper, Nicholas Humphrey (1976; see also Jolly 1966) published an important paper in which he argued for what is now called the Machiavellian intelligence hypothesis, which we briefly discussed above. The hypothesis holds that the high level of intelligence and relative brain size observed in primates (particularly, chimpanzees) is best explained as a result of the unique demands that their social living places on them. To survive and thrive in their complex social world, Humphrey argued, primates needed to evolve certain mindreading abilities to solve complex social problems. Inspired by Humphrey's hypothesis, a number of field researchers came forward with reports of anecdotal cases of primate intentional deception (or 'tactical deception') that appeared

to show the use of innovative behavioral strategies by these animals for the purpose of inducing false beliefs in conspecifics. Many of these early field observations of intentional deception were collected and discussed in Byrne and Whiten (1988) and Whiten and Byrne (1997).[7] One such case deserves special mention for its subsequent influence in the field.

The American primatologist Emil Menzel (1974) was conducting spatial memory experiments on a group of captive chimpanzees when he observed some rather novel behavior in one of his adult females, called Belle. In the study, Menzel showed Belle the location of some hidden food in the 1-acre field where the chimpanzees lived and then returned Belle to her group mates. The group was then released into the field. On every occasion of their release, Belle made a straight line for the hidden food. The alpha male of the group, Rock, eventually caught on to this pattern of behavior and began to follow Belle to the hiding place of the food, whereupon he would quickly push her aside and consume all the food. On one occasion of the group's release, however, Belle unexpectedly went in the opposite direction from where she saw the food hidden. Rock, quite predictably, followed close behind. However, while Rock was preoccupied with looking in the wrong place, Belle quickly doubled back and consumed the hidden food.

Menzel and others took the novelty and apparent ingenuity of Belle's behavior as evidence of an explicit intention to deceive Rock, an intention to induce a false belief about the location of the food. In the early 1990s, Byrne and Whiten (1990) compiled a large database of similar observations of intentional deception in primates by field researchers. Eighteen of these accounts were identified by the researchers as cases of intentional tactical deception among great apes (Byrne 1995). More recently, researchers have run tests similar to Menzel's with chimpanzees (Hirata & Matsuzawa 2001) and mangabey monkeys (Coussi-Korbel 1994). Both studies report intentional deception in their animals similar to that observed in Belle.

At around the time of the release of Byrne and Whiten's database, Dorothy Cheney and Robert Seyfarth (1990) began to report the results of their experiments on the alarm-calling behaviors of wild vervet monkeys in Africa. Although prior to this, vervets had been known for their repertoire of distinct alarm calls for different types of predators (e.g., calls for eagles, calls for snakes, calls for neighboring troops of vervets, etc.), and for their flexible and appropriate responses to the distinct calls (e.g., looking up in the sky upon hearing an eagle alarm call, looking into the bushes upon hearing a python alarm call, etc.), much was still unknown

about the intentional and semantic aspects of their communicative behaviors. Through detailed observations and a series of ingenious playback experiments using hidden speakers, Cheney and Seyfarth discovered that vervets not only appeared to possess some voluntary control over their calls (e.g., calling only when other vervets were known to be nearby) but they appeared to understand the different meanings or semantic information carried by the different alarm calls. From a hidden speaker in the trees, Cheney and Seyfarth played a recorded alarm call (e.g., a martial eagle alarm call) from an individual monkey at a time when the monkey was out of sight from the rest of its troop. Cheney and Seyfarth observed that the other members of the troop quickly habituate (i.e., stop responding) to this individual's alarm call if (after repeated playbacks) it was shown to be unreliable (e.g., when no martial eagle was ever seen). What is more, the researchers discovered that the troop also stopped responding to other semantically related but acoustically distinct alarm calls from this same individual (e.g., other raptor alarm calls). Quite remarkably, though, the troop continued to respond as usual to this individual monkey's semantically unrelated calls (e.g., leopard alarm calls) as well as to the semantically identical calls (e.g., martial eagle calls) from different members of the troop. The monkeys, it seemed, were evaluating the reliability of the individual's calls not on the basis of the calls' brute acoustical properties but in terms of the information the calls carried. Although Cheney and Seyfarth were reluctant to interpret their findings as proof that these vervet monkeys were attributing communicative intentions (a type of mental state) to callers, their findings did suggest to some that the monkeys were engaging in a form of intentional communication that may involve a rudimentary form of (and perhaps an evolutionary precursor to) such mental state attribution (see Gómez 2009).

There were critics of these various field studies, however. Some were skeptical of the data on the grounds of its anecdotal nature (Premack 1988; Povinelli 1996; Heyes 1998), while others were skeptical on more substantive grounds. During the mid and late 1990s, there was a growing skepticism among some researchers about the possibility of mindreading in animals. Chief among these skeptics was Daniel Povinelli and his colleagues. Povinelli and Eddy (1996) ran a series of discrimination tasks with chimpanzees to test their understanding of the mental state of seeing in others, and they received across-the-board negative results from them. In the first phase of the experiment, the chimpanzees were trained to beg for food from a single trainer. They were then tested with two new trainers, one who could see the chimpanzee and one who could not. In some

of the test trials, for example, the seeing trainer faced forward while the unseeing trainer faced backward, while in other trials, the seeing trainer wore a blindfold around his neck while the unseeing trainer wore a blindfold over his eyes. The chimpanzees showed no signs of an initial preference for begging from the seeing trainer over the unseeing one. They did, however, show some improvement over time, preferring to beg from the seeing trainer more often than from the unseeing trainer. But the chimpanzees' incremental success, Povinelli and Eddy argued, was best explained in terms of their coming to follow a simple rule of thumb (e.g., pick the trainer whose face is visible) learned during the test trials which had nothing to do with their understanding the psychological state of seeing.[8]

A couple of years later, Cecilia Heyes (1998) published an important critical review of all extant data on primate mindreading and forcefully argued that none of it singly or collectively made a compelling case. Three significant points were made in the article. First, Heyes stressed, as Harman, Dennett, and Bennett had done twenty years earlier, the critical importance of overcoming the logical problem. To make progress on the question of animal mindreading, Heyes argued, researchers needed to design tests that could adequately distinguish genuine mindreading in animals from various plausible behavior-reading capacities in them. Second, Heyes showed in some detail that no experimental approach at that time was capable of solving the logical problem. And finally, and most importantly, she proposed an alternative experimental paradigm, the experience-projection (EP) paradigm, which she argued could effectively discriminate genuine mental state attribution from various forms of behavior reading in animals.

In the version of the EP paradigm that Heyes described in her paper, a chimpanzee is allowed to discover something about its own mental state of seeing while wearing (for the first time) two different kinds of goggles. One pair of goggles (trimmed in blue) was fitted with a clear lens that would allow the chimpanzee to see objects in the environment; the other pair of goggles (trimmed in red), however, was fitted with an opaque lens that would prevent the chimpanzee from seeing objects in the environment. It was speculated that the chimpanzee might learn, through wearing the different goggles, that it could see objects in the environment while wearing the blue-trimmed goggles but not while wearing the red-trimmed ones.

After its exposure to the goggles, the chimpanzee would then be tested to see if it would use this knowledge about its own mental states of seeing/

not seeing to discriminate between a seeing trainer (wearing the transparent blue-trimmed goggles) and an nonseeing trainer (wearing the opaque red-trimmed goggles) in a knower–guesser protocol similar to that used by Premack (1988). Heyes reasoned that if the chimpanzee was a mindreader that was capable of introspecting its own states of seeing, it would be expected to choose the container indicated by the trainer with the blue-trimmed goggles on the grounds that that trainer, rather than the one wearing red-trimmed goggles, likely *saw* which container was baited. However, if the chimpanzee were but a behavior reader, and could appeal only to the observable, mind-independent facts presented in the pretest trials, then, Heyes reasoned, it should show no such bias in its choice of containers in the test trials since neither the blue- nor the red-trimmed goggles were ever associated with a container having food inside it.

To this day, Heyes's experimental protocol with goggles has not been run on animals.[9] Inspired by Heyes's protocol, however, Emery and Clayton (2008) recently ran a modified version of it (sans goggles) with scrub jays, the results of which, they argue, show that these birds are capable of projecting their own experience of pilfering caches onto conspecifics who are currently observing them caching. Emery and Clayton's experiment will be discussed in detail in chapter 2. Notwithstanding its infrequent use in animal mindreading research, the EP paradigm has come to be seen by some researchers (Povinelli & Vonk 2006; Lurz 2009a) as the only experimental method that can solve the logical problem. I shall make a strong case for this view in chapter 3.

Research into animal mindreading lay relatively dormant during the late 1990s, perhaps due to the negative findings and critical review discussed above. Then, at the beginning of the new millennium, the tide dramatically changed. Brian Hare and colleagues (2000, 2001) ran a series of unique mindreading experiments on chimpanzees that involved competition rather than cooperation, as in the traditional knower–guesser paradigm. In Hare and colleagues' study, a subordinate chimpanzee and a dominant chimpanzee competed for food placed behind different types of barriers. The researchers discovered that subordinate chimpanzees consistently refrained from taking food that was out in the open or behind clear (transparent) barriers—that is, food that the dominant chimpanzee in the situation could see—but attempted to take food that was behind opaque barriers—that is, food that the dominant chimpanzee in the situation could not see. The researchers argued that the subordinate chimpanzees' performance could not easily be accounted for in terms of learned associations or simple behavioral rules of thumb but was best explained by credit-

ing the chimpanzees with an understanding of the difference between seeing and not seeing in dominant conspecifics. The results have since been replicated and found to be robust (Bräuer et al. 2007; however, see Karin-D'Arcy & Povinelli 2002), leading a number of researchers to believe that chimpanzees are capable of attributing states of seeing to others.

After the negative results from Povinelli and Eddy's experiments involving cooperation between chimpanzee and human trainers, the positive findings from Hare and colleagues' experiments suggested to a number of researchers that the key to unlocking mindreading capacities in chimpanzees and other animals was competition (Hare 2001; Hare & Tomasello 2004; Lyons & Santos 2006; Santos et al. 2007). The competitive paradigm is currently the dominant approach used to test for mindreading in animals. To date, it has been used on monkeys (Flombaum & Santos 2005; Santos et al. 2006), scrub jays (Emery & Clayton 2008), ravens (Bugnyar & Heinrich 2005), bee-eaters (Watve et al. 2002), and goats (Kaminski et al. 2006), all with positive results.

At the same time that Hare and colleagues introduced the competitive protocol to behavioral studies of animal mindreading, a group of scientists and a philosopher were making the case for investigating mindreading in animals at the neuronal level. Back in the mid 1990s, a group of Italian neuroscientists discovered a cluster of neurons (F5 neurons) in the premotor cortex of monkey brains that fired not only when the monkeys were about to perform an intentional action, such as grasping an object, but when they observed another subject (monkey or human) performing the same type of intentional action (Gallese et al. 1996; Rizzolatti et al. 1996). In watching another subject's action, the monkey's own brain appeared to mirror (or resonated with) the observed subject's own premotor intention to act. The monkey, it was argued, in observing the other's action, came to share the intention (though, without acting on it) that it took to be behind the other agent's observed action. Hence the F5 neurons were aptly labeled "mirror neurons." At the same time as this discovery, philosopher Alvin Goldman was defending a radically new theory of mindreading called the simulation theory. The simulation theory viewed mindreading as essentially a process by which a creature uses its own mind as a model or replica of another's mind in order to determine which mental states to attribute, much in the same way that one might use a scaled model to make predictions about or attribute properties to a real object. It did not take long before the two groups came together to advance the view that F5 neurons in monkeys enable these animals to attribute the mental states of goals and intentions to others by a process of mental simulation

(Gallese & Goldman 1998; Gallese et al. 2004; Fogassi et al. 2005; Goldman 2006).

At roughly the same time, researchers in the behavioral tradition were engaged in a number of new studies that probed goal-directed/intentional-action attribution in monkeys and apes. Call et al. (2004) reported evidence that chimpanzees spontaneously (i.e., without prior training) discriminate between a trainer who failed to share food because he was unable but *willing* to share and one who failed because he was *unwilling* but able to share, and Wood et al. (2007) have recently reported evidence that both monkeys and apes discriminate between rationally directed and irrationally directed actions in others.

Despite the positive findings from these new mindreading experiments, some have remained skeptical. Povinelli and colleagues have argued that none of the new experimental approaches fares any better than the old ones at solving the logical problem. All the positive data from these new types of experiments, they maintain, can just as easily be explained in terms of behavior-reading capacities in animals. They argue further that the new experimental approaches cannot even in principle distinguish mindreading from various behavior-reading strategies and are therefore utterly useless for answering the question of whether animals are capable of mindreading. In addition, Povinelli and colleagues (Vonk & Povinelli 2006; Penn et al. 2008) have recently advanced a general theory about the difference between human and animal minds, called the reinterpretation hypothesis. According to this hypothesis, various sorts of cognition in humans, including mindreading, involve reinterpreting first-order perceptual facts and behavioral patterns in terms of nonobservable higher-order causal relations. Animals, Povinelli and colleagues argue, lack concepts for nonobservable higher-order causal relations and are therefore incapable of mindreading. We shall examine this argument and others made by Povinelli and colleagues in chapter 3.

As we saw above, philosophers have made some important contributions to the animal mindreading debate. Some of their views will be explored in more depth in chapter 4. However, two deserve special mention here for their importance in shaping the history of the debate. Arguably the best known and most influential of these philosophers is Daniel Dennett. In a series of papers dating back to the late 1970s (reprinted in Dennett 1987), Dennett proposed and defended the intentional stance strategy. The basic idea behind the intentional stance strategy, as it bears on the question of mindreading in animals, is that there is nothing more to animals' being mindreaders (what Dennett calls "second-order inten-

tional systems") than the fact that researchers find it useful and economical to predict and explain the observed patterns of the animals' behaviors by means of mindreading (second-order intentional) hypotheses. There are no deeper facts about animals' brains or the structure of their internal representations, on this account, that are relevant to whether they attribute mental states. The intentional stance strategy, thus, takes a decidedly instrumentalist (i.e., anti-realist) view of what it is for animals to be mindreaders.

The instrumentalism of the intentional stance strategy is nicely illustrated by the strategy's account of the injured-wing display of piping plovers. These birds, on seeing a likely predator approaching their nest, are known to move away from their nests and then, in full view of the predator, to hop around on the ground with one wing outstretched. This display typically causes the predator to move away from the nest in pursuit of the apparently injured bird. Before being captured, the plover flies off and continues the display until the predator is a safe distance from the nest (Ristau 1991). Regarding this pattern of apparently intelligent and deceptive behavior in piping plovers, Dennett (1978b) writes,

[W]hen we ask why birds evolved with this tropism, we explain it by noting the utility of having a means of deceiving predators, or inducing false beliefs in them; what must be explained is the provenance of the bird's second-order intentions [i.e., the intention to cause in predators the belief that its wing is injured].… [W]hat I want to insist on here is that if one is prepared to adopt the intentional stance without qualms as a tool in predicting and explaining behavior, the bird is as much a second-order intentional system [i.e., a mindreader] as any man. Since this is so, we should be particularly suspicious of the argument I was tempted to use, viz., that representations of second-order intentions would depend somehow on language. For it is far from clear that all or even any of the beliefs and other intentions of an intentional system need be represented "within" the system in any way for us to get a purchase on predicting its behavior by ascribing such intentions to it. (pp. 276–277)

Piping plovers, on the intentional stance strategy, are thus genuine mindreaders—they attribute beliefs to predators—even though there is nothing in their brains, no higher-order representation, that represents predators as having beliefs or any mental state at all.

Unsurprisingly, researchers who take up a more realist interpretation of the mind tend to object to Dennett's account of what mindreading in animals amounts to. To some, the intentional stance strategy, as a result of it rejection of the idea that mental state attributions in animals are or involve higher-order representations in their brains, is destined to make

mindreading in the animal kingdom too commonplace to be considered empirically plausible or interesting. Lowly mantis shrimp, after all, are also known for their 'feigned' threat displays while molting (Adams & Caldwell 1990), but few (if any) researchers have ever thought to conclude from this that these stomatopods are as much mindreaders as any human. Their reluctance, undoubtedly, is due to the fact the brains of these animals are just too simple to support such complex, higher-order cognition. In addition, researchers have worried that the intentional stance strategy would unduly rob their mindreading hypotheses from offering causal explanations of animal behavior. As the plover example illustrates, there need be nothing in an individual animal's brain or body that is the animal's belief that another creature has such-and-such mental state for the animal to be said to have such a higher-order belief. Thus, on the intentional stance strategy, the fact that an animal has a belief about another creature's mental state cannot be assumed to be what *caused* the animal to anticipate the creature behaving as if it had such a mental state (on the assumption that the causes of animals' anticipatory behaviors are facts or events that occur within their brains or bodies). For these reasons, some scientists and philosophers have eschewed Dennett's intentional stance strategy for more realist accounts of what mindreading in animals entails.[10]

One such philosopher is José Luis Bermúdez. Bermúdez (2003, 2009) has forcefully argued for just the sort of deeper fact that Dennett's intentional stance strategy denies is relevant to the question of animal mindreading. On Bermúdez's account, mindreading about propositional attitudes (e.g., beliefs and desires) involves processing internal representations that are sentences in the subject's own public language. Without the possession of a public language to think in, Bermúdez argues, animals are incapable of attributing mental states with propositional content and are, thus, restricted to attributing (if at all) simple perceptual and motivational states that lack propositional content. This restriction holds, according to Bermúdez, even if (pace Dennett) researchers find it abundantly useful and economical to interpret animals as attributors of beliefs and desires, or any other propositional attitude. Bermúdez's view will be examined further in chapter 4.

1.3   Conclusion

Obviously much has been left out of this short history of the debate, but enough has been provided, I believe, to reveal in outline the main lines of disagreement. Chief among these is the logical problem—the problem of

how to determine by experimental means whether an animal uses a representation of another agent's mental state (e.g., the agent *sees* the food) to anticipate its behavior or just the behavioral/environmental cues associated with the mental state in the other agent (e.g., the agent has a unobstructed line of gaze to the food). In the following chapters, we shall examine the exact nature of this problem more closely, as well as whether it has been, needs to be, or even can be solved.