“Starting from what I know of the operations of my own individual mind, and the activities which in my own organism they prompt, I proceed by analogy to infer from the observable activities of other organisms what are the mental operations that underlie them.” So wrote John George Romanes (1882, pp. 1-2), outlining his method for gaining scientific leverage on the problem of animal minds. In the wake of Darwin’s (1871/1982) publication of *The Descent of Man*, Romanes took up the challenge of investigating the evolution of mind with a fervor. Darwin had marshaled an impressive array of observations to suggest that not only were humans descended from other species in bodily structures, but also in mental structures. For Romanes, the agenda for a new science was clear. Just as anatomists “aim at a scientific comparison of the bodily structures of organisms,” he observed, “so [comparative psychology] aims at a similar comparison of their mental states” (Romanes, 1883, p. 5). But Romanes faced a problem. Although his analogy between comparative anatomy and a new science of comparative psychology was powerful, it began to break down when it came to the substances to be compared. Anatomists had access to dead bodies, but the stuff of psychology was not so easily available for examination on the laboratory bench. Recognizing this, Romanes offered an interim solution by turning to the only source of material available—the spontaneous behavior of animals. But even here, he was far behind the anatomist in that there was no existing corpus of data. Romanes knew he might be roundly chastised for doing so—and he was—but ultimately he was forced to rely on anecdotes as his database. “If the present work is read without reference to its ultimate object of supplying facts for
the subsequent deduction of principles," Romanes apologized, "it may well seem but a small improvement upon the works of the anecdote-mongers" (1882, p. vii).

Although the foundations of Romanes' approach collided with an age-old philosophical problem far more profound than that of inferring mental states in other species—the uncertainty of making inferences even about the minds of our fellow humans—in a very real sense, the core of his method was unassailable. After all, we are not just bodies cohabiting the same physical locations. We are thinking, feeling beings, negotiating a social world teeming with each others' desires, goals, intentions, and emotions. We are linked not just in space, but also in mind. Each of us possesses an intense desire to have our lives understood by others. Indeed, this desire to understand and be understood, to be part of a group—whether it be the village, school, family, or kibbutz—would appear to be as universal a trait of the human species as any.1 This inextricable connection between self and other even occurs in our most silent moments. Someone looks around a corner, furrows his or her brow, and we effortlessly attribute a visual experience of having just seen something. In this sense, Romanes' solution to the problem of understanding animal minds was simple: If we can make reasonably accurate inferences about what other humans are thinking and feeling by just observing them, why can we not do the same with members of other species? If the method of introspective analogy can bridge the gap that separates the mind of one human from another, why cannot it not likewise span the distance that separates one species from another?

In this chapter, we return to Romanes' problem by examining the evolution of the psychological connection between self and other that appears so characteristic of our species. We explore three related questions. First, when did the ability to conceive of the self and others as mentalistic agents evolve? That is, when and in what lineages did organisms first evolve the capacities to reason about themselves and others in terms of mental states and events such as thinking, knowing, believing, attending, desiring, and perceiving? The data we review suggest that humans may be largely unique in being able to reason about such internal states. Second, we ask whether the evolution of self—other psychology occurred in synchrony, or whether an understanding of self as a mental agent preceded a comparable understanding of others. A main issue we address is whether the fusion of self and other understanding we observe in humans can be dissociated in other species. For example, the data we review are consistent with the hypothesis that although chimpanzees may possess at least a limited objective self-concept, they may lack the ability to conceive of others (and perhaps even themselves) as mentalistic agents. Finally, we address the thorny problem of how it is that nonhuman primates can share with us so many behavioral
patterns that in humans are clearly associated with a mentalistic understanding of others, yet not possess such understanding themselves. Our conclusion is that humans may have evolved these abilities not because they endowed us with scores of novel behaviors per se, but rather because they allowed for the complex reorganization and redeployment of existing behavioral patterns. Thus, although humans may indeed possess unique behavioral capacities (e.g., active pedagogy), the initial utility of interpreting behavior in terms of internal, nonobservable mental states may have been at an organizational level—at a higher level of abstraction than any specific set of behaviors. If this general idea is correct, it may mean that traditional efforts to find a coherence between an understanding of self and others as mental agents, and some set of naturally occurring behaviors, will remain largely futile.

In order to set the stage for examining these ideas, we review several aspects of current theorizing and research concerning self and social understanding. First, by examining certain aspects of the spontaneous social behavior of primates in nature, we explore how the traditional approach to understanding the evolution of social intelligence has led to the idea that many, if not most, nonhuman primates possess some understanding of the mental states of themselves and others. Second, we explore the evidence concerning the evolution of self-conception and explain why these data have been interpreted as suggesting a qualitative psychological difference between the great apes and humans on the one hand, and most other forms of life on the other. Third, we examine experimental data concerning the development of one aspect of human infants and children’s understanding of their own and others’ minds, and parallel research with chimpanzees. Our interpretation of these data is that despite their striking similarity to us at the level of specific behavioral patterns, not even chimpanzees possess a theory-of-mind system comparable to that which develops in humans during late infancy and early childhood.

**SELF, SOCIALITY, AND INTELLIGENCE**

**Evolving an Intelligence of Others**

Alison Jolly (1966) authored the first careful statement of the idea that the truly remarkable features of primate intelligence had evolved in the context of coping with each other—a kind of social intelligence of others. Returning from an early field survey of the prosimian primates of Madagascar, Jolly was left with the impression that primate intelligence was far from a unitary concept. She reflected that intelligence about objects and physical events appeared to have evolved independently from intelligence deployed
in the social realm. She minimally differentiated between intelligences "toward objects, including food; toward other active species, including predators; and toward fellow members of one's own species" (p. 504). Some of her own previous laboratory research had suggested that prosimians lagged far behind other primates both quantitatively and qualitatively on standardized tests of intelligence (Jolly, 1964a, 1964b; see also Andrew, 1962). In contrast, the sophistication of their social behaviors seemed comparable to what had been previously observed in various monkey species, including the characteristically long chains of social interactions.

Given that prosimians appeared able to generate the same level of social complexity as the anthropoid primates, but were far less proficient at traditional object-oriented laboratory tests of intelligence, it occurred to Jolly that prosimians might share a different form of intelligence with other primates—a social intelligence. As she explained:

The social use of intelligence is of crucial importance to all social primates. As the young develop, they depend on the troop for protection and for instruction in their role in life. Since their dependence on the troop both demands social learning and makes it possible, social integration and intelligence probably evolved together, reinforcing each other in an ever-increasing spiral. And although it is very likely that the learned social relations of monkeys are in fact more complex than those of lemurs, our present techniques of description emphasize the similarity between lemur and monkey social interactions. (1966, p. 504)

Jolly's speculations amounted to suggesting a possible dichotomy or compartmentalization between these two types of intelligence. Indeed, experimental evidence from rhesus monkeys reared in social isolation was already hinting that there might be some validity to this claim by demonstrating that despite their inability to learn a variety of species-typical patterns of social behaviors, they performed like normal monkeys on most traditional, object-oriented intelligence tests (Harlow, 1965; Harlow, Schlitz, & Harlow, 1968).

Nicholas Humphrey (1976) independently reached a similar conclusion in the mid-1970s. Like Jolly, he noted that most primates, on the surface, appeared to possess a veritable "surplus" of intelligence: a seemingly unneeded cornucopia of intellectual abilities unrelated to the demands of their way of life in nature. Given his view of nature as an uncompromising optimizer, Humphrey quickly reached the conclusion that the idea that such a surplus truly existed "was most likely to be wrong" (p. 303). The solution, he reasoned, might lie in the function that intellect played in the social arena:

[The] social primates are required by the very nature of the system they create and maintain to be calculating beings; they must be able to calcu-
late the consequences of their own behaviour, to calculate the likely beh-
aviour of others, to calculate the balance of advantage and loss—and all of this in a context where the evidence on which their calculations are
based is ephemeral and liable to change, not least as a consequence of
their own actions. (p. 309)

And, also like Jolly, he took special note of the self-generating nature of in-
tellect in this social context. Social intelligence was driven through an evol-
utionary ratchet effect, “acting like a self-winding watch to increase the
general intellectual standing of the species” (p. 311).

But what about the interconnection of self and other—did the social
intelligence hypothesis (as the Jolly–Humphrey idea came to be known)
have anything to say about the apparent fusion of these intelligences? In
later essays, Humphrey (1980, 1982) traced out the implications of his
ideas for precisely this problem. Marshaling a long-standing philosophical
position, he argued that once self-consciousness had provided our species
with a way of making introspective sense of our own behavior, then our
own experiences and ways of understanding behavior would “immediately
and naturally [be] project[ed] onto other people” (Humphrey, 1982, p.
477). Thus, like others before him, Humphrey concluded that humans
come to understand the inner thoughts and desires of those around them
through an introspective examination of their own mental states and
processes—a kind of mental simulation of what it must be like to be the
other person (for variations on this solution to the problem of other minds
see Hume, 1739; Smith, 1759/1961; Adams, 1928; Russell, 1948; Stewart,

Humphrey’s speculations about the evolution of the connection be-
tween self and social understanding were largely confined to humans—but
were other species capable of generating inferences about the mental lives
of those around them? Humphrey (1982) offered cautious skepticism on
this point: “I am not yet convinced that any other species has followed the
same path to consciousness as man…” (p. 477). But, to be fair, he noted
that “[i]t may turn out there are, in fact, non-human species…[that are]
making use of explanatory systems which bear the hallmarks of a mind cap-
able of looking in on the inner workings of the brain” (p. 477). As we
shall see, at least one other comparative psychologist hypothesized that
chimpanzees and some other great apes might be in this very same episte-
mological position (see Gallup, 1982, 1983).

Social Manipulation and Deception in Primates

If Jolly and Humphrey were right, and the social dimensions of intellect
had been under intense selection due to the burgeoning sociality of the pri-
mates, then surely there must be examples of such complex uses of this in-
telligence in the day-to-day interactions of these animals. And indeed, by the late 1960s, the evidence for such maneuverings was already starting to surface. After a lengthy field study of hamadryas baboons, Kummer (1967) reported instances of females presenting and glancing at a dominant male while threatening a rival female. He speculated that this behavior functioned to manipulate the male into attacking the other female. By the mid-1980s, Smuts (1985) was urging primatologists to recreate the common sense of their discipline around emotional and mental states such as ambivalence, flirtation, trust, jealously, affection, and grief. For example, she offered preliminary evidence for a system of “selective retaliation” in savanna baboons by showing that after baboons had an aggressive encounter with more dominant animals, the animals apparently waited to retaliate—surreptitiously and selectively—against the relatively subordinate “friends” of the animals that had initially aggressed against them.

Meanwhile, other researchers had been reporting even more striking episodes of social manipulation—episodes that appeared to qualify as instances of intentional deception (e.g., Menzel, 1974). In Tanzania, for example, Jane Goodall (1971) reported several cases of chimpanzees acting in ways that seemed to suggest that they were capable of suppressing normal behaviors in order to mislead each other. A decade later, Frans de Waal’s (1982) Chimpanzee Politics—a careful blend of quantitative data and anecdotal accounts—helped to establish chimpanzees’ reputation as Machiavellian beasts par excellence (see also, de Waal, 1986). But chimpanzees were not to claim exclusive rights to such status for long, as evidence of deception in other primate (and nonprimate) species was soon to follow. Byrne and Whiten (1985) described examples of apparent deception among familiar baboons. They distinguished between deception well known to ethologists (such as mimicry and camouflage) and what they dubbed “tactical deception.” They defined tactical deception as those instances in which an individual used an act from its normal behavioral repertoire in a unique context, so that it served to manipulate another individual. For instance, in one of the episodes they reported, a juvenile male baboon, who was being approached by threatening males, suddenly stopped, stood on his hind legs, and looked into the distance. This gesture was familiar to the team of researchers—it was a signal typically given in the context of an approaching predator or another group of baboons. The other males, apparently reacting to the warning, abandoned their hostile ambitions toward the juvenile and scanned the horizon as well. But when the human observer scanned the horizon, there was no trace of a predator or rival baboon troop. After several years of scouring the literature and soliciting unpublished accounts from other primatologists, Whiten and Byrne (1988) concluded that many different species of primates engage in tactical
deception. In formalizing Humphrey’s (1980) claim that primates are natural psychologists, Whiten and Byrne were attempting to revive the anecdotal method as a means of gaining insight into the social psychology of nonhuman primates.

As we have seen, Romanes (1882) had also attempted a careful use of anecdotes to infer the kinds of mental states other organisms must be experiencing. Likewise, after categorizing their examples of deception, Whiten and Byrne (1988) aimed to use these anecdotes to “sketch the features of the state of mind that an individual with deceptive intent must be able to represent” (p. 233). Whiten and Byrne noted that their database of anecdotes could never be more than a starting point for “more systematic” work (p. 243). But by “more systematic” they were not necessarily calling for experimental approaches to the question. Instead, they were directly advocating the collection of more detailed observations in natural settings. But could Whiten and Byrne’s rebirth of the anecdotal method succeed where Romanes (1882, 1883) had failed? Even more generally, could the spontaneous behavior of free-ranging primates tell us anything at all about whether these animals interpret each other in terms of unobservable mental states?

The skeptics were un convinced. One set of problems had to do with the ontogenesis of the behaviors. “The plural of anecdote,” Irwin Bernstein (1988) quipped, “is not data” (p. 247). Gordon Burghardt (1988) agreed: “How does [Whiten and Byrne’s] call for multiple records really counter the problems inherent in interpreting any ‘single observation’? How much problematic data adds up to one conclusive bit of evidence?” (p. 249). David Premack (1988) was more blunt: “How many ‘trials’ go into producing the anecdotes that are reported from the field? Since this is rarely known, readers are led to indulge their ignorance and to draw romantic conclusions” (p. 171). But a second set of problems hinged on the more fundamental issue of whether, regardless of their ontogenetic origins, such behavior must be based upon a mentalistic interpretive system. Premack put the problem simply: “In calling [an] observation an anecdote, I do not mean to question the reliability of the report or even, for that matter, the general accuracy of the circumstance in which the act occurs” (p. 162). But if many, highly reliable anecdotes could be coded and catalogued, why would the need for experimentation exist? “The need for experiments,” Premack replied, “arises with respect to the interpretation we wish to place on the act” (p. 162). Thus, in the view of Premack and other skeptics, reports of the spontaneous actions of nonhuman primates, as intriguing as they are, bring us no closer to answering the fundamental question posed by Edward Thorndike’s discovery of the law of effect over a century ago: Do animals understand social interactions in terms of behavior alone, or both behavior and mind?
Self-Recognition, Self-Conception, and Theory of Mind

While field researchers were observing the spontaneous behavior of nonhuman primates and speculating about the evolution of self and social understanding, an experimentally minded comparative psychologist, Gordon G. Gallup, Jr., was also reflecting on the relationship between the evolution of self-consciousness and the ability to reason about the minds of others. In a series of ingenious experiments conducted in the late 1960s, Gallup (1970) had uncovered a peculiar quirk of psychological evolution: Humans, chimpanzees, and presumably other great apes, were capable of recognizing themselves in mirrors, whereas various species of monkeys he tested were not. Gallup reported that after several days of mirror exposure, chimpanzees spontaneously learned to use mirrors to explore parts of themselves that they had never (or only rarely) had the opportunity to see before. Thus, although they initially reacted to the mirror as if they had been suddenly confronted with another chimpanzee, his subjects soon altered their disposition toward their images and began using the mirror to manipulate and explore their ears, eyes, noses, teeth, and anogenital areas—all while carefully monitoring the effects in the mirror (see Figure 3.1).

In order to provide more direct experimental evidence for his impression that the subjects had learned to recognize themselves, Gallup anesthetized the chimpanzees and marked them with a bright red, odorless, tactile-free dye on their upper eyebrow ridge and ear. These areas were selected because in absence of a mirror, the apes would not know that parts of their faces had been dyed red. After recovery from the anesthesia, the subjects were observed for 30 minutes in the absence of the mirror, and virtually never touched the marks. In direct contrast, and in clear support for the idea that they had learned to recognize themselves, when the mirror was reintroduced, the chimpanzees reached up to touch these otherwise invisible marks. Various aspects of Gallup's findings with chimpanzees have been replicated both within this species, and in orangutans, but generally not gorillas2 (Lethmate & Ducker, 1973; Ledbetter & Basen, 1982; Suarez & Gallup, 1981; Swartz & Evans, 1991; Calhoun & Thompson, 1988; Povinelli, Rulf, Landau, & Bierschwale, 1993; Eddy, Gallup, & Povinelli, 1996). More recent research has pointed to striking developmental and individual variability in the phenomenon (Swartz & Evans, 1991; Povinelli et al., 1993; Eddy et al., 1996).

Heyes (1994, 1995) has recently attempted to discount both the spontaneous reactions of chimpanzees to mirrors, as well as the results of the mark tests as evidence of self-recognition. However, these alternative explanations have not fared well against either theoretical or empirical scrutiny (see Figure 3.2A-3.2B; Gallup et al., 1995; Eddy et al., 1996; Povinelli et al., 1997). Although not all great apes display the ability to recognize
FIGURE 3.1. Chimpanzees show evidence of recognizing themselves in mirrors. In contrast to other species, many chimpanzees engage in self-exploratory behaviors (C–F). They also use mirrors to make and view exaggerated facial displays (A and B). Many chimpanzees that display the types of self-exploratory behaviors that are shown in C–F also will pass a “mark test” (through use of a mirror, these animals often display emphasized inspection of parts of their faces which have been surreptitiously marked red).

themselves in mirrors (especially gorillas), results from our laboratory, using both cross-sectional and longitudinal designs and several new methodological techniques, have converged to support the idea that by about 5–8 years of age, many chimpanzees display unambiguous evidence of using mirrors to gain information about their physical appearance (Povinelli et
al., 1993, 1997; Eddy et al., 1996). Indeed, the experimental techniques used with chimpanzees have allowed a more sophisticated diagnosis of the presence of mirror-mediated, mark-directed behavior in chimpanzees than in 18- to 24-month-old children (see Figure 3.2A, B).

Coincident with his discovery that chimpanzees are capable of recognizing themselves in mirrors, Gallup (1970) also uncovered an additional finding that many comparative psychologists have found difficult to accept: the apparent absence of self-recognition in primate species outside the great ape/human group. Despite weeks, months, and in some cases a lifetime of exposure to their mirror images, researchers working with other primate species have been unable to demonstrate the pattern of phenomenon reported for chimpanzees and orangutans (Gallup, 1970, 1977; Leth- mate & Ducker, 1973; Benhar, Carlton, & Samuel, 1975; Gallup, Wallnau, & Suarez, 1980; Anderson, 1983; Bayart & Anderson, 1985; Suarez & Gallup, 1986; Itakura, 1987; Anderson & Roeder, 1989; Fornasieri, Roeder, & Anderson, 1991; Gallup & Suarez, 1991). Attempts to demonstrate self-recognition in a wide range of nonprimate species (including elephants and dolphins) have thus far not produced compelling positive evidence (Povinelli, 1989; Marino, Reiss, & Gallup, 1994; for a review of the reactions of a variety of species to their mirror images, see Gallup, 1968, 1975). This is not to say that in the quarter-century that has elapsed since Gallup's original report appeared there have been no claims for self-recognition in one or more monkeys, and even dolphins. To be sure, there have been several (e.g., Boccia, 1994; Thompson & Boatright-Horowitz, 1994; Marten & Psarakos, 1994; Hauser, Kralik, Botto-Mahan, Garrett, & Oser, 1995). However, the best available evidence continues to point to a qualitative functional difference between the responses of many great apes versus other primates to their mirror image (see Gallup, 1994, for a review).

But what about the theoretical significance of these findings? Gallup (1970) saved the final two paragraphs of his original report to account for the phylogenetic difference he had apparently uncovered: “Such a decisive difference between monkeys and chimps,” he struggled, “is particularly interesting in view of the fact that most investigators have found only relatively slight quantitative differences on other, more traditional behavioral tasks” (p. 87). His pivotal point was summarized in a single, guarded comment: “Insofar as self-recognition of one’s mirror-image implies a concept of self, these data would seem to qualify as the first experimental demonstration of a self-concept in a sub-human form” (p. 87). Gallup concluded that the data suggested that a qualitative psychological difference among primates had been detected, and that it seemed doubtful that the capacity for self-recognition extended “below [sic] man and the great apes” (p. 87).

As efforts to facilitate self-recognition in monkeys continued to yield negative results, and as the number of innovative techniques designed to
FIGURE 3.2. (A) An advanced experimental technique used to assess mark-directed behavior compares touches to comparable marked (filled areas) and unmarked regions (open areas) on a subject's head. Raters score videotapes for touches to each of these four regions in 30-minute control period (mirror covered) and 30-minute test period (mirror uncovered). (B) Duration (top panel) and frequency (bottom panel) of touches to marked and unmarked facial regions. Subjects selectively touch marked regions in test period (Povinelli et al., 1997). From Povinelli et al. (1997), Copyright 1997 by Academic Press Limited. Reprinted by permission.
rule out simple procedural explanations mounted, Gallup continued to ponder the potential significance of this apparent qualitative difference between monkeys and apes. It struck Gallup that chimpanzees must possess some form of self-awareness. In short, he reasoned that an organism must first know that it exists before it can recognize a reflection of itself (Gallup, 1975, 1977, 1979). If his chimpanzees had no self-concept, he argued, then there would have been no way for them to have known that the marks were on themselves—and hence no reason to search the corresponding areas on their own bodies (see Davis, 1989, for a philosophical defense of Gallup’s inference). As Gallup (1977) put it,

The unique feature of mirror-image stimulation is that the identity of the observer and his reflection are necessarily one and the same. The capacity to correctly infer the identity of the reflection must, therefore, presuppose an already existent identity on the part of the organism making this inference. . . . While many organisms are ostensibly conscious of different features of themselves as the result of visual, tactile, chemical, and proprioceptive feedback, in principle this is quite different from self-consciousness. (p. 334)

Evidently, Gallup reasoned, his chimpanzees had not only grasped the duality that is apparent in any mirrored surface, but they also possessed a sufficiently well-integrated self-concept to realize that the dualism applied to themselves as well. Indeed, Gallup (1982) later speculated even further. The presence of a self-concept in chimpanzees, he suggested, might reflect a limited capacity to introspect about some subset of their own mental experiences.

An organism’s view of itself was one thing, but how would such a creature view other organisms? Gallup (1992) reasoned that self-aware organisms might understand others in much the same fashion as they understand themselves. If they explain their own behavior in terms of emotions, desires, intentions, and beliefs, perhaps they might assume that others have similar mental states as well. Such organisms, like humans, might use their own experiences to make inferences about what other creatures in similar situations might think and feel. Thus, Gallup proposed that evidence of self-recognition in mirrors—because of its possible connection to the presence of self-awareness—might be indirect evidence for an ability to understand the minds of others. His idea, therefore, united self and social intelligence, and in particular, intelligence related to the ability to introspect about one’s own mental states (Gallup’s self-awareness) and the ability to make inferences about analogous mental states in others.

Gallup’s ideas constituted a theory in that they generated what were, at the time, some rather striking predictions about both those species that
had displayed evidence of self-recognition in mirrors and those that did not. The theory predicted that whereas the former would succeed on tests designed to tap into their ability to reason about the private, unobservable mental states of others, the latter would not. Indeed, if the common ancestor of the great apes and humans had uniquely evolved a psychological system related to the self-concept—a system that simultaneously allowed them to make inferences about the mental states of others—then there ought to be a class of identifiable social behaviors unique to its living descendants as well (e.g., Gallup, 1982, 1983, 1985). In other words, Gallup (1982) speculated that if the great apes possessed a psychological system that distinguished them from other primates, then there ought to be corresponding behavioral differences as well. To illustrate his argument, Gallup quickly surveyed the primate literature, pulling out examples of empathy, deception, grudging, and the like, to illuminate his ideas. His intent was not to conduct an exhaustive review of the literature on primate social cognition, but rather to highlight a class of behaviors that could make his comparisons explicit. Of course, he realized that some researchers might immediately reject the notion of a dichotomy between great apes and other nonhuman primates because of widespread anecdotes of deceptive, empathic, and other behaviors that would, at first glance, appear to qualify as evidence of reasoning about mental states.

But Gallup drew a more subtle distinction between self-aware instances of these behaviors, on the one hand, and learned or hardwired (innate) ones on the other. As an example, he drew on Gene Sackett’s (1966) demonstration that very early in their development, isolation-reared infant rhesus monkeys show different (and appropriate) reactions to photographs of adult male rhesus monkeys who display threatening faces and those that do not—despite the fact that they have never had the social interactions in which they could have learned this distinction. Gallup explained the significance of such findings:

In the absence of ever having had any experience with adult males there would be no basis upon which to make that distinction. From an adaptive point of view, however, it is not important that [the infants] know, but only that they act as if they knew something about the apparent intentions of threatening adults males. (Gallup, 1982, p. 245)

Thus, nature was far too complex to simply “read into” animals an understanding of intentions, desires, and beliefs, as Romanes had counseled a century earlier. Instead, clear standards were needed for distinguishing between real and learned/hardwired instances of mind. Indeed, as primatologists interested in the mental states of animals began to rely on anecdotes (especially those concerning deception) more and more frequently, Gallup’s
suggestions for how to distinguish among such cases began to take a different turn. In particular, as his concerns about the difficulties in separating learned/hardwired behaviors from those based on mental state attribution mounted, he began to favor more controlled experimental demonstrations of these abilities (compare Gallup, 1982, 1985, 1991). Indeed, he found it increasingly necessary to explain why the interaction of learning and hard-wired developmental mechanisms on the one hand, and anthropomorphic tendencies in humans on the other, meant that naturalistic observations—no matter how reliable—were not well-suited to answer questions of this type. "Mindlessness," Gallup (1985) commented in reference to the illusions of mind that can be produced through these mechanisms, "is not obvious" (p. 634).

As an alternative to using anecdotes, Gallup turned to David Premack and Guy Woodruff’s (1978) report concerning the possibility that chimpanzees possessed a theory of mind. By “theory of mind” they were referring to the effortless way in which humans make inferences about the mental states of themselves and others. “A system of inferences of this kind,” they observed, “may properly be viewed as a theory because such states are not directly observable, and the system can be used to generate predictions about the behaviors of others” (p. 515). Turning to Premack’s star pupil, Sarah, Premack and Woodruff attempted to determine if chimpanzees could reason about the intentions of others. They constructed a series of videotape sequences that depicted familiar humans struggling to solve staged problems, and then gave Sarah ample opportunity to view them. Each videotape was then played for her, placed on pause, and she was provided with pairs of still photographs—one of which represented the solution to the implied problem, the other, which did not. So, for example, on one trial, Sarah witnessed an actor struggling to obtain a banana suspended from the ceiling out of reach. Sarah was presented with two photographs, one in which the actor was stepping onto a box, the other in which the actor was engaging in an irrelevant activity. Sarah’s above-chance success in selecting the photographs that depicted the solution to a range of such problems was interpreted as evidence that she “recognized the videotape as a problem, understood the actor’s purpose, and chose alternatives compatible with that purpose” (p. 515). Many researchers have challenged Premack and Woodruff’s interpretation of their data, but suffice it to say that their conceptualization of the problem of the attribution of mental states, and their experimental approach to the question, impressed Gallup (1982, 1983) as a straightforward means of testing a central prediction of his model: that organisms that displayed clear-cut evidence of recognizing themselves in mirrors ought to be capable of reasoning about the mental lives of others.

On the other hand, Gallup’s theory predicted that species incapable of
self-recognition did not possess an introspective system, and hence could not use their own experiences to simulate or model the experiences of others. Likewise, in a later developmental application of these ideas, Gallup and Suarez (1986) argued that if the theory was correct, young children should not begin to show signs of self-awareness or of using their own experiences to make sense of the behavior of others until they displayed evidence of recognizing themselves in mirrors. Indeed, subsequent research provided evidence for such correlations with respect to the development of empathy, altruistic behaviors, the self-conscious emotions, and synchronic play (Asendorpf & Baudonniere, 1993; Lewis, Sullivan, Stanger, & Weiss, 1989; Brownell & Carriger, 1990; Johnson, 1982; Bischof-Köhler, 1988, 1991).

Gallup’s hypothesis provided a formalism to the idea that primate evolution had been marked by a connection between self and social understanding, and offered a set of specific, testable predictions concerning the evolutionary history of this connection. Like Humphrey (1982), Gallup argued that our indirect, inferential knowledge of others is derived from our direct, primary access to our own mental and emotional states. In this sense, Gallup’s proposal was striking not so much for its claims about how knowledge about the minds of others is obtained, but rather for its startling claim that self-recognition in mirrors, because of its purported relationship to introspection, might be a useful phylogenetic (and developmental) marker of the presence of the ability to know such things about others in the first place. What Gallup’s model did not do—indeed, did not purport to do—was to provide an explanation for why these forms of self-awareness had evolved in the first place.

**Self-Recognition and Self-Conception: Alternative Views**

Gallup’s model, while making some clear predictions regarding mental-state reasoning in different primate species, left open the question of the exact nature and scope of the self-representation that allows for self-recognition in mirrors in the first place (see Gallup, 1979; Povinelli, 1991; Mitchell, 1993). Gallup’s (1970) initial speculation was that “self-directed and mark-directed behaviors would seem to require the ability to project, as it were, proprioceptive information and kinesthetic feedback onto the reflected visual image so as to coordinate the appropriate visually guided movements via the mirror” (p. 87). It was only later, as Gallup began to consider the connection between self and other, that he hypothesized an even more elaborate form of self-awareness in chimpanzees—one that included an ability to introspect about one’s own mental states (Gallup, 1982, 1983, 1985). It is important to note that this later formulation was not offered as deductively valid: Nothing forced the conclusion that chim-
panzees were self-aware in this latter sense. Rather, Gallup offered a speculative model that contained clear and falsifiable predictions.

Several investigators have proposed alternatives to, or elaborations of, Gallup’s ideas about the cause of self-recognition in mirrors (Epstein, Lanza, & Skinner, 1981; Mitchell, 1993; Parker, 1991). Although they differ in a number of ways, most of these models either explicitly or implicitly assume that before organisms can display evidence of self-recognition in mirrors, they must first learn that mirrors provide contingent accurate images of objects in front of them. This line of thinking has led researchers to approach the mark test as a problem in which the organism is searching for the correct location of the red marks they see in the mirror (see Bertenthal & Fischer, 1978; Mitchell, 1993).

We have recently proposed an alternative to this line of thinking about the nature of the self-representation underwriting self-recognition in mirrors (Povinelli, 1995). Although our model shares much in common with Gallup’s (1970) original view, and Mitchell’s (1993) subsequent elaboration of some of Gallup’s ideas (Mitchell’s “inductive” model), our explanation begins by denying that self-recognition in mirrors has anything to do with understanding the reflective property of mirrors; that is, we have argued that understanding that mirrors provide an accurate and contingent image of one’s own body is not necessary for mirror self-recognition. Indeed, we suspect that despite their capacity to recognize themselves in mirrors, not even 18-month-old human infants know this fact (see Povinelli, 1995).

How, then, is self-recognition in mirrors possible? A majority of human infants display evidence of self-recognition in mirrors (and other forms of live visual feedback) by about 18–24 months (Amsterdam, 1972; Schulman & Kaplowitz, 1977; Bertenthal & Fischer, 1978; Lewis & Brooks-Gunn, 1979; Johnson, 1983; review by Brooks-Gunn & Lewis, 1984). Because far more is known about aspects of cognitive development in human infants, we start by asking about other changes that occur at around 18–24 months of age that might support this ability. Povinelli (1995) has argued that the typical development of a domain-general system of representation in 18- to 24-month-old children provides them a means of forming relations between objects and events in the world and schemes held in mind. Although they differ in important ways, a number of specific proposals are compatible with this view (e.g., Perner, 1991; Olson & Campbell, 1993). Once such a system of representation becomes possible, an organism is able to construct a number of first order relations, but of particular interest to us is their ability to grasp equivalence relations. Elsewhere, we have outlined how the ability to form such equivalence relations allows the organism to understand that the representation of self held in mind (an immediate, on-line representation of the self’s actions, move-
ments, and possibly desires) is equivalent to the image they see in the mirror (see Povinelli, 1995). Everything that they can detect about the self's appearance, movements, and desires (as recognized in action) is equivalent to the image in the mirror. This does not require the organism to understand that the mirror is an accurate and contingent image of itself. This fact may or may not be understood, but it is not necessary for self-exploratory and mark-directed activities to occur. Following William James (1890/1950), we have labeled this self-representation the "present self."

One of the central components of our model is that it posits the contingency between the self's actions and the actions in the mirror, is what triggers the formation of an equivalence relation between the organism's internal self-representation and the external stimuli (the mirror image). Thus, if 18-month-old children were able to understand not only the relation between their present self and live visual feedback, but also understood how their present self is related to previous states of the self, our model would be in serious jeopardy. Recently, we have tested whether such a developmental dissociation exists in young children's capacity for self-recognition using contingent visual feedback (e.g., mirrors, live video) and delayed visual feedback (e.g., prerecorded video). In a series of studies, we have shown that although 2- and 3-year-old children will reach up to remove a sticker that was secretly placed on their head if they see themselves in a mirror or in live video, they will not typically do so if they see a delayed videotape that depicts the experimenter placing the sticker on their head 3 minutes earlier (Povinelli, Landau, & Perilloux, 1996). It is not that the children fail to recognize their physical features on the delayed tape; to the contrary, they are quite willing to use their proper name or the first-person pronoun "me" when asked who it is (see also, Brooks-Gunn & Lewis, 1984). Rather, they do not seem to understand that the image has anything to do with their current self—their on-line representation of their actions and desires. Indeed, consistent with the theoretical predictions of our model, it is not until 4–5 years of age that children pass this test of self-recognition using delayed feedback (see also Povinelli & Simon, 1998).

These empirical findings have important implications for assessing the scope of the chimpanzee's self-concept. Although to date we have only conducted pilot tests with chimpanzees using our delayed self-recognition procedures, our theoretical model explicitly argues that organisms such as chimpanzees can recognize themselves in mirrors without necessarily appreciating that they are unique, unduplicated selves with a particular past and possible futures. Passing our tests of delayed self-recognition, in contrast, requires a representational system that allows them to link particular instances of the self into a higher-order self that endures through time. Indeed, there appears to be a relationship between these phenomena and the development of autobiographical memory in children. Our findings of the
age at which young children appear to grasp the connection between their current state and briefly delayed video images of themselves, mirrors the age of onset of autobiographical memory (see Nelson, 1992, 1993; Welch-Ross, 1995). Although at present we do not yet know whether chimpanzees develop an ability to pass tests of delayed self-recognition, their ability to recognize themselves in mirrors uniquely counts only as evidence for a limited self-concept restricted to their on-line actions and possibly desires. One curious fact remains, however. As we have seen, the emergence of self-recognition in mirrors in human infants is significantly correlated with the emergence of other abilities that may relate to an understanding of other minds (e.g., empathy, altruism, synchronic play, the self-conscious emotions). If similar correlations held true for apes, Gallup's model would turn out to have heuristic merit. However, as we shall see, such correlations may not be present in chimpanzees.

**THEORY OF MIND AND THE SELF–OTHER FUSION**

In contrast to the simulation view adopted by Humphrey and Gallup, a very different relation between self and other has recently been advanced by the "theory" theorists. Following Premack and Woodruff's (1978) lead, the theory-theorists have argued that like adults, young children (and possibly even infants) possess a naive, folk "theory of mind" that is used to explain and predict the behavior of others (Wellman, 1990; Perner, 1991; Gopnik & Wellman, 1994; Gopnik & Meltzoff, 1997). These researchers use the term "theory" quite literally, meaning that our beliefs about other minds are acquired, revised, and used in the very same manner as are scientific theories—it is just not as formalized (see Gopnik & Meltzoff, 1996). Thus, in characterizing the development of young children's knowledge about mental states, these researchers argue that the child really and truly is constructing a bona fide (if naive) theory of how the mind works. The development of knowledge about the mind, in their view, can best be understood as a series of revisions of theories about how the mind operates—revisions forced by the accumulation of new evidence (Wellman, 1990; Gopnik & Wellman, 1994; Gopnik & Meltzoff, 1997).

The theory-theory has direct implications for the means by which we come to know the minds of those around us. Alison Gopnik (1993), for example, has maintained that the intuitive, cherished belief that we have direct access to our own mental states, but only indirect access to the mental states of others, is simply wrong (Gopnik, 1993; Gopnik & Wellman, 1994; Gopnik & Meltzoff, 1997). Thus, in direct contrast to the position adopted by Humphrey, Gallup, as well as by Paul Harris (1991) and Robert Gordon (1986), Gopnik has argued that the way in which we know
about the mental states of others is the very same way in which we know about our own. She has proposed that from early infancy forward, knowledge of the mental states of self and other are derived from the same psychological system, and in particular, a system that forms theories about how the social world operates. But, just as our knowledge of the mental states of others is indirect and inferential, so too, she claims, is our knowledge of our own mental states. As we develop, we invent and learn about theoretical mental constructs such as goals, desires, intentions, and beliefs that guide our interpretation of both our own behavior and that of others. Thus, Gopnik maintains that mental states are every bit as theoretical when we invoke them to explain our own behavior as when we use them to explain the behavior of others.

The relation between self and other minds proposed by Gopnik (1993) is certainly a strong version of the claim that knowledge of self and other are inextricably connected. Indeed, not all developmental psychologists concur with her characterization of the self–other relation, or with her more general allegiance to the idea that human infants and children are forming “theories” about how the world operates. Although their reasons for doing so vary widely, there are many who find it impossible or difficult to accept the idea that young infants and children actively construct “theories” of mind during their development (e.g., Leslie, 1994; Fodor, 1987, 1992; Hobson, 1993). However, it is important to note that even if we were ultimately to reject the extreme position of the theory-theorists, we would still be left acknowledging that humans possess an uncanny psychological connection to each other. In a particularly rich attempt to understand the developmental origins of our psychological connection to others, Peter Hobson (1993) argues that this psychological identification has “pretheoretical origins,” and argues that the earliest manifestations of understanding the psychological attitudes of self and other occur by observing mental states that, in some sense, are directly observable. But placing the particular developmental accounts aside for the moment, an important phylogenetic question remains: Are humans alone in this psychological fusion of self and other? More broadly, are humans alone in possessing a theory of mind?

In the sections that follow, we select a particular case study of theory-of-mind development in humans and examine what is known about comparable abilities in chimpanzees. In particular, we explore their understanding of seeing as a mental event. Ideally, our review would be more broad, including a wide range of abilities associated with the development of theory of mind in human children, and would include data from not just chimpanzees but other nonhuman primates as well. Unfortunately, at present, such a review is impossible given that there have been very few systematic investigations of theory-of-mind abilities in nonhuman primates. Even
in the case of chimpanzees, where the most research has been conducted, the kinds of systematic studies that will be necessary to come to closure on these issues are in their infancy (see Tomasello & Call, 1994; Povinelli, 1996a). Nonetheless, by comparing chimpanzees' and children's understanding of seeing as a projection of attention, we hope to shed some light on both the similarities and differences in the cognitive systems of human and nonhuman primates. We hope to show how it is possible that the evolutionary emergence of theory-of-mind abilities forever altered the manner in which humans understand the behaviors they share in common with their primate ancestors.

Some might question the logic of comparing the psychological development of human children and other species, such as chimpanzees, on any number of grounds. Elsewhere, we have explored these objections and have explained the methodological and biological rationale of such comparisons in detail (see Povinelli & deBlois, 1992a; Povinelli & Eddy, 1996a; Premack & Dasser, 1991). In summary, despite the separate evolutionary histories of these species, comparisons of the type we review here can help to establish features of psychological functioning that these species inherited from a common ancestor, as well as unique features that one of more of the species have evolved after their lineage diverged from that ancestor. In short, systematic comparisons of chimpanzee and child, despite their methodological difficulty, can ultimately provide a fair characterization of the unique cognitive abilities of the two species.

DEVELOPING A THEORY OF MIND:
THE CASE OF SEEING-AS-ATTENTION

Before we begin our examination of what is currently known about the development of young children's understanding of the act of seeing as the mental state of attention, it is important to emphasize that this research is really just a small subset of a much broader field of inquiry. Since Premack and Woodruff's (1978) report, investigations of the development of young children's knowledge about mental states have escalated at an unprecedented rate. Currently, researchers are investigating young children's understanding of a wide range of mental states and events such as consciousness, desire, attention, intention, knowledge, belief, false belief, and thinking. A number of developmental progressions have been offered to describe the child's developing theory of mind (Wellman, 1990; Perner, 1991; Gopnik & Meltzoff, 1996). One of the most widely accepted is a scheme offered by Henry Wellman (1990), in which young children are postulated to move from desire–belief reasoners in the second year of life to belief–desire reasoners from about 3 or 4 years on. He argues that by about 2–3 years of
When Self Met Other

age, young children conceive of the mental affairs of others primarily on the basis of desires, although they can and do reason about the beliefs of others to a certain extent. In contrast, he believes that by about 4 or 5 years of age, children have firmly grasped the role that beliefs play in regulating the behavior of others, and use this understanding as the centerpiece of their reasoning about the mental life of others. Other researchers have argued that even prior to this desire-belief reasoning, human infants possess a kind of desire-goal psychology in which they understand the action of those around them in terms of observable goal states that are internally desired (Baron-Cohen, 1991). Indeed, others hold out for a continuous extension of intersubjective experiences into early infancy (Gopnik & Meltzoff, 1997).

Several researchers have attempted to provide a unified account of the development of the child's knowledge about the mind. One approach has been to ignore the messy problem of development altogether by postulating that cognitive development is really just the onset of brain modules, prewired to perform the neural computations necessary to reason about the phenomena in question (Fodor, 1983, 1992; Leslie, 1994; Baron-Cohen, 1994, 1995). A second approach has been to specify domain-general representational systems that the child constructs during development and then to show how changes in these general structures allow for the behaviors identified at various ages (Perner, 1991; Olson, 1993; Frye, Zelazo, & Palfai, 1995; Zelazo & Frye, in press; Gopnik & Meltzoff, 1997).

Although there are clear theoretical points of disagreement between and among these approaches, it is important not to overlook two fundamental points of unity. First, they all start with the notion that humans do, at some point, construct a psychological system that allows them to reason about unobservable mental states of the self and other. Second, they agree that however difficult the task, if the relevant environmental parameters and the transformational processes that act upon them were completely understood, we would be able to characterize the modal developmental progression of the developing child's knowledge about the mind. As we shall see, these seemingly trivial points of agreement have nontrivial implications. We now examine a specific case in point concerning the child's growing understanding of the mind—the development of their understanding of seeing-as-attention.

Understanding Seeing: From Meaningful Stimuli to Mental State

Many organisms—not just primates—are interested in and sensitive to the presence of eyes and eye-like stimuli (e.g., see Blest, 1957; Burger, Gochfeld, & Murray, 1991; Burghardt & Greene, 1988; Gallup, Nash, &
Ellison, 1971; Ristau, 1991; Perrett et al., 1990; Povinelli & Eddy, 1996a, 1996b, 1996c, in 1997; see review by Argyle & Cook, 1976). A moment’s reflection on the impact of predation on both solitary and group-living organisms reveals why the process of natural selection may have favored the preservation of perceptual systems that reacted quickly to the presence and direction of the eyes of other organisms. In the case of group-living organisms, such as many primate species, additional advantages to paying attention to who is looking at whom (or what) can easily be envisioned (Chance, 1967; Fehr & Exline, 1987). In many of these cases, we are not readily tempted to conclude that the organisms in question know anything at all about a psychological state of attention behind the eyes—for example, the eye-like stimuli on the wings of certain moths and butterflies (Blest, 1957). But other situations readily tempt our anthropomorphic inclinations. For example, surely a rhesus monkey who leaps up to threaten you after you look her in the eyes must understand that you looked at her. Surely a chimpanzee who spins around to look where you are looking, must have understood that you were looking at something behind him. Well, maybe—but maybe not.

Let us assume for the moment that some organisms can and do process information about the eyes of other organisms without having any notion at all about seeing. Thus, just as many birds have specific courtship displays in which a behavioral act by one of the participants reliably triggers a response by the other one, so too may the reactions of many species to the eyes and eye direction of others have evolved as useful responses to specific stimuli. So, although many species actively monitor and respond appropriately to the visual attention of others, they may not be aware that others have visual experiences. They simply may not interpret visual perception as a mental event. We have not yet presented any evidence that they do not; rather, we are simply holding open this possibility.

On the other hand, we can easily imagine that some species do understand visual perception as an intentional event linking the perceiver to the external world. By “intentional” we refer to the philosophical notion of “aboutness”—that is, seeing is about (or refers to) organisms, objects, or events in the external world (Brentano, 1874/1960). Indeed, we do not merely have to imagine such species, for we know of at least one in which this kind of understanding is well consolidated: our own. Indeed, the eyes seem to occupy center stage in our human folk psychology. The commonly used metaphor of making “eye contact” with someone highlights this property well; although our eyes do not literally contact the other person, our folk psychology foists this connection upon the interaction. When we see someone turn to look at an object, we automatically register that they are attending to that object.
Developing the Idea That Seeing Is Attention

Are humans born with this mentalistic understanding of the eyes, or does it develop gradually? Even 4- to 5-month-old infants are sensitive to the presence of eyes and will look longer at faces that make direct eye contact with them as opposed to those that do not (Lasky & Klein, 1979; Johnson & Vicera, 1993). Between 6 and 18 months, infants construct an ability to follow the gaze of others by turning to look in the same direction (Scaife & Bruner, 1975; Butterworth & Cochran, 1980; Butterworth & Jarrett, 1991; Corkum & Moore, 1994). George Butterworth and his colleagues have proposed three stages in the development of this ability. Initially, infants only appear able to turn their heads in the same direction as the other person. They limit their visual scans to the space that they directly perceive but cannot yet localize the specific object to which the adult is gazing. Later, by about 12 months, this ability is elaborated, allowing infants to localize the specific target of the adult’s gaze, again, provided it is within their own visual field. Finally, by about 18 months or so, infants will turn and search in space outside their own immediate visual field if, for example, the adult glances behind them.

Some researchers interpret gaze following as evidence of an intentional understanding of seeing—a kind of joint visual attention in which the infant and the adult are aware of each other’s attentional focus on a particular object or event in the external world. This interpretation implies that an infant’s ability to follow another person’s line of sight expresses an understanding that the person is looking “at” something (Baron-Cohen, 1994, 1995; Franco, in press). Other theorists, although impressed by the sophistication of the gaze-following system, are far less sanguine about its relation to an intentional understanding of visual perception (Butterworth & Jarrett, 1991; Moore, 1994; Tomasello, 1995; Povinelli & Eddy, 1994, 1996a, 1996b). Simply because infants are interested in the eyes, and can use the gaze direction of others to discover useful information in the world, this does not guarantee they appreciate that seeing subjectively connects the observer to the external world. Thus, although a gaze-following system ensures that mother and infant will look at the same object in unison, it is far less clear that the infant is aware of this fact. In this sense, joint attention does not ensure shared attention.

If sensitivity to the presence and direction of eyes does not necessarily qualify as evidence that infants understand the underlying attentional significance of visual perception, then what might? First of all, it is important to clarify that infants might develop a general understanding of attention before understanding the exact role that the eyes play in regulating its deployment and direction. Generally, when we attend to something, there are
multiple redundant cues indicating our attentional focus, including eye direction, facial orientation, and bodily posture. Indeed, these are merely the cues that concern visual attention. Often, we attend to things by listening, touching, or smelling. Thus, an infant might construct the general notion of attention as a mental state long before realizing either that there are specific sensory channels that each offer distinct input, or the factors that govern the deployment of each channel. For example, infants might conclude that people can be attending to an object if their head and body are oriented toward it—even if their eyes are closed! Thus, we must ask two questions: (1) At what point do infants come to appreciate the mental state of attention, and (2) at what point do infants appreciate the eyes in particular as portals through which attention is regulated?

With respect to the first question, there is evidence that by about 18 months (and possibly earlier), infants may be able to represent the mental state of attention. For example, Dare Baldwin and her colleagues have used several procedures in which infants are asked to reason about the attentional focus of an adult. In one of them, an experimenter shook an opaque plastic bucket to demonstrate that something was inside. Next, the experimenter looked into the bucket without letting the infant see, and declared, "It's a modi! A modi! There's a modi in here!" Next, the experimenter picked up an identical bucket, shook it, and then looked inside and pulled out a toy that the child had never seen before and let the child play with it. Finally, the experimenter pulled a comparable, unfamiliar toy from the first bucket and let the child play with it as well. The crucial test came after both toys had been retrieved from the infant and then placed in front of him or her. The experimenter said, "There's a modi here. Can you point to the modi? Point to the modi." Of course, seeing the person peer into the bucket would automatically indicate to us that he or she was talking about whatever was inside. Baldwin's (1993a) results suggest that by about 19 months or so, young children are apparently in this same boat. They correctly select the object that the experimenter was presumably looking at while labeling the object. Numerous control procedures have been conducted to rule out alternative explanations, and, in general, the data have converged to suggest that somewhere in the middle of their second year of life, infants are able to interpret the actions of others in terms of a hypothetical mental state that we call attention (Baldwin, 1991, 1993a, 1993b; Baldwin & Moses, 1994; Mumme, 1993). Whether infants younger than 18 months have a similar, if more circumscribed and fragile, competence in this area remains a matter of controversy (see Baldwin & Moses, 1994).

But what about the eyes? When can we feel confident that infants have uncovered the connection between the eyes in particular, and the mental state of attention? That is, when do they understand seeing-as-attention? In an impressive series of studies spanning over a decade, John
Flavell and his colleagues have asked precisely this question (Masangkay et al., 1974; Lempers, Flavell, & Flavell, 1977; Flavell, Shipstead, & Croft, 1978; Flavell, Flavell, Green, & Wilcox, 1980; Flavell, Everett, Croft, & Flavell, 1981; Flavell, Green, & Flavell, 1989). However, they have distinguished two levels at which seeing can be understood mentalistically (see Figure 3.3A). At Level 1, young children may understand that seeing connects people to the external world; in other words, they can register what people can see and what they cannot see. Their studies have revealed that by about 2½ years of age, young children appear to have localized the eyes as being relevant for seeing. So, for example, if you ask young 3-year-olds to make it so that you cannot see a ball, they know to move a screen between you and it; if you ask them to show the ball to someone else who is facing away, they will hold the object up and place it in the other person's

**FIGURE 3.3.** (A) By about 2½ years of age, children understand that seeing connects people to the external world. For example, they understand that people can see or not see based on the direction of their gaze. (B) In addition, by 4 years of age, children develop an understanding of how visual perception leads to the formation of internal states of knowledge.
line of sight; if you stare at one of several toys and ask them to point to the one you can see, they do so with ease; if you cover your eyes, they know you cannot see them (see Lempers et al., 1977). Povinelli and Eddy (1996a) and Gopnik, Meltzoff, and Esterly (1995) have provided additional evidence for this kind of understanding of seeing-as-attention in 2½- and 3-year-old children. Finally, using cartoon stimuli of faces, Baron-Cohen and his colleagues have shown that 3-year-old children can discern what another person wants based solely on the direction in which the character's eyes are directed (Baron-Cohen, Campbell, Karmiloff-Smith, Grant, & Walker, 1995).³

When Seeing Becomes Believing (and Knowing)

Although 2- and 3-year-olds seem to understand visual perception as a mental state, it would be erroneous to conclude that they have achieved an adult-like understanding of seeing. Indeed, it is likely that their understanding of seeing is quite shallow. Flavell and his colleagues have identified a second level of understanding seeing that children of this age do not yet grasp. Although they have no trouble realizing what you can or cannot see, they have striking difficulty understanding how those things appear to you. For example, sit across the table from a young 3-year-old girl and show her a picture of a turtle so that it is right-side-up from her perspective, but upside-down from yours. She will readily assert that you can see the turtle, and indeed, if you cover your eyes, she will readily acknowledge that you can no longer see it. But try as you may, you will have an extraordinarily difficult time getting her to understand that, from your perspective, the turtle appears differently, that is, upside-down (e.g., Flavell et al., 1981). But less than a year later, this same child will demonstrate without hesitation her understanding that although the two of you are visually connected (or attending) to the same thing, your respective mental representations of the object differ considerably (Level 2 understanding, Figure 3.3B). Although one might be tempted to dismiss these findings as being the result of some trivial artifact of the experimental design, Flavell and his colleagues have used procedure after procedure to assist the younger children, generally with little or no success.

Indeed, Flavell's findings seem to tightly parallel other differences discovered between 3- and 4-year-olds related to their understanding of the connection between seeing and knowing. For example, 3-year-olds have great difficulty in understanding the link that perception in general (and seeing in particular) plays in the process of knowledge formation (Wimmer, Hogrefe, & Perner, 1988; Perner & Ogden, 1988; Gopnik & Graf, 1988; Ruffman & Olson, 1989; O'Neill & Gopnik, 1991; Povinelli & deBlois, 1992b; O’Neill, Astington, & Flavell, 1992). When 4- or 5-year-olds see
When Self Mer Other

someone lift a lid on a box and look inside, they automatically assume that the other person knows what is there. In contrast, they do not make this assumption if someone just touches the box, or lifts the lid without looking inside. Even though they cannot literally see information entering the person’s eyes, traveling to the brain, and resulting in a particular activation of neurons that correspond to the state of knowing the contents of the box, these children appear to assume that internal states of belief have arisen from the act of visual perception (see Figure 3.3B). Yet if the same scenario is rehearsed for 3-year-olds, it rapidly becomes apparent that they are interpreting these events in a very different manner. They appear to have little idea that the one who sees, knows, and that the one who does not see, does not know.

Indeed, 3-year-olds’ difficulty is not limited to their understanding of visual information. Their difficulty appears to be a general one in that they have not yet understood that auditory, visual, haptic, and olfactory information lead to, or cause, internal knowledge states. Of course, there is some disagreement on this point, with some researchers offering evidence that 3-year-olds, for example, may possess some understanding of the seeing–knowing relation (e.g., Pillow, 1989; Pratt & Bryant, 1990; Wooley & Wellman, 1993; but see critiques by Povinelli & deBlois, 1992b). However, Lyon (1993) has recently provided evidence that the disagreement could stem from the fact that younger children may conflate the idea of knowledge with interest or desire. His research suggests that young 3-year-olds attribute “knowledge” to those people who display an interest in the contents of a box, more than to those who look inside the box, but express little other evidence of interest. O’Neill (1996) has also recently provided evidence that 2½-year-olds are sensitive to whether their mothers have been visually or attentionally engaged with an event, and if they have not been, will attempt to draw their mothers’ attention to the situation. However, she has argued that despite this sensitivity to their mothers’ behavioral connection or engagement to a given event, children of this age may still possess little or no appreciation of knowledge states per se.

In parallel to their trouble in understanding the perception–knowledge relation in others, 3-year-olds have great difficulty in explaining how they themselves come to know things. For instance, let a young 3-year-old look into a bag and discover a toy bird inside, then put the bag away. A minute later, if you ask him if he remembers what was inside the bag, he will answer correctly. But then ask him how he knows there is a bird in the bag. Here, he will look at you blankly—just before rapidly describing all sorts of irrelevant (if true) details about the dog he has at home that one time tried to bite the mailman. You can rephrase the question, make it as obvious as possible for him, and even give him a choice: “Do you know there’s a bird in there because I told you or because you looked inside?” Try as
you may, you will discover that young preschoolers appear to have little idea how, or even when, they came to know such a simple fact (Wimmer et al., 1988; Gopnik & Graf, 1988; O’Neill & Gopnik, 1991; O’Neill et al., 1992; Povinelli & deBlois, 1992b; Taylor, Esbensen, Bonnie, & Bennett, 1994).

Summary: Seeing as a Mental Event

The findings reviewed here suggest that during the earliest months of life, children are sensitive to the head posture, face, and eyes, and by the end of the first year have consolidated a simple gaze-following system that allows them to track where someone else is looking (as long as it is within their own immediate perceptual field). Furthermore, by 18 months or so, children appear to have constructed a mentalistic notion of attention, even if they have not yet sorted out the exact role that the eyes and other sensory modalities play in mediating that attention. By 2½ or 3 years of age, preschoolers seem to be well on the way to understanding the eyes in particular as portals through which attention emanates. Finally, by about 4 or 5 years of age, children come to understand how perception brings information into the mind; that is, they come to understand the role that seeing (along with the other senses) plays in creating the mental states of knowledge and belief.

As we noted earlier, a number of authors have discussed how some of these facts can be explained in more general terms either by implicating changes in a domain-general representational system, the activation of specific brain modules, or the progressive revision of naive theories of how the mind works (Wellman, 1990; Perner, 1991; Leslie, 1994; Olson, 1993; Baron-Cohen, 1994, 1995; Frye et al., 1995; Zelazo & Frye, in press; Gopnik & Meltzoff, 1996). We recognize that not everyone will find our summary in keeping with the majority of the data. However, even if specific proposals about the exact timing of children’s development of an understanding of seeing (and their theory of mind in general) are rejected in the face of new data, the general theoretical possibilities we outline should remain tenable.

COMPARATIVE EVIDENCE: WHAT CHIMPANZEES KNOW ABOUT ATTENTION

Chimpanzees: Gaze Following

Until recently, there were only anecdotal reports that chimpanzees and other nonhuman primates follow each other’s line of sight (e.g., Byrne &
When Self Met Other

Whiten, 1985). Indeed, this evidence, combined with a generous “joint-attentional” interpretation of the behavior, led some investigators to conclude that chimpanzees possess an understanding of the intentional aspect of seeing (e.g., Baron-Cohen, 1994). But do chimpanzees (and other non-human primates) really follow gaze, and if so, what is the extent of similarity between the gaze-following system that is elaborated between 6 and 18 months in human infants and any similar system in other species?

Over the past several years, our laboratory has conducted a series of studies designed to explore the gaze-following phenomenon in 7 preadolescent chimpanzees. The results of these experiments have uncovered striking similarities between human infants and chimpanzees. First, chimpanzees appear to display gaze following in response to movement of the head and eyes in concert, or just the eyes alone (Povinelli & Eddy, 1996b, Experiment 1; see Figures 3.4 and 3.5). Second, like 18-month-old human infants (but not younger ones), chimpanzees will track the gaze of an experimenter into regions of space not within their immediate perceptual field (Povinelli & Eddy, 1996b). So, if you gaze above and behind a chimpanzee, there is a very good chance he or she will wind up looking there almost immediately. Third, chimpanzees appear to extract specific information about the direction of gaze. Thus, looking behind them does not simply trigger a general visual scanning response on their part; if you look to their right, they will look there first as well (Povinelli & Eddy, 1997; Povinelli, Bierschwale, &

FIGURE 3.4. Like 18-month-old human infants, chimpanzees follow the gaze of an experimenter into regions of space not within their immediate perceptual field. In (A), the chimpanzee enters the test room. In (B), the experimenter turns her head to a region in space not within the immediate visual field of the chimpanzee subject. The subject responds (C) by rapidly orienting head and body toward the same general direction as the gaze of the experimenter.
Cech, in press, Experiments 1 and 2). Finally, chimpanzees do not even need to witness the shift in another person's line of sight in order to engage in gaze following (Povinelli & Eddy, 1996a, Experiment 12). Rather, if they merely encounter you with your head and eyes in a fixed orientation, directed above and behind them, this is sufficient to cause them to turn and look where you are looking. People who witness this behavior for the first time are generally deeply impressed. However, not everyone is equally convinced that this behavior qualifies as evidence of an appreciation of the mental state of attention. As we noted in discussing the development of gaze following in human infants, there are good reasons to consider alternative, lower-level explanations of the underlying causes of any such behavior.

In order to gain some empirical leverage on this problem, we attempted to challenge our chimpanzees' gaze-following system in a way that could reveal more about their understanding of gaze. A low-level account posited that they were just automatically registering the sudden shift, or unusual orientation of the eyes/face of others, and then turning along a particular trajectory until their visual system oriented to a novel object or event. A high-level model assumed that they interpreted our line of sight as a projection of attention (see Figure 3.6). We subjected these competing
FIGURE 3.6. (A) A prototypical gaze-following event in which the animal on the left first sees an event (a descending raptor), and an observer (on the right) utilizes the first animal's gaze to discover the event. Solid lines represent the line of sight of the gazing animal; dash-and-dot lines represent the line of sight of the animal tracking the gaze. This utilization of gaze may be based on a variety of mechanisms, some being more mechanically based and others being more psychologically based. A low-level model posits that the observer follows the gaze of the first animal without any understanding of the mental state of attention. Rather, it uses a system that projects a trajectory from the face of the first animal out into distant space. Alternatively, a high level model posits that the observer recognizes the change in the first animal's mental state of attention, and this information prompts the gaze-following response. (B) Some naturally occurring situations might assist in distinguishing between the two accounts for a given species. With the animal on the left seeing an event behind a bush, the animal on the right might use the first animal's gaze in one of two ways. The low-level model predicts that without an understanding of attention per se, the "gaze" could be projected directly through the bush, and past it. In contrast, the high-level model expects that the observing animal could ascertain that the first animal must be looking at something on its own side of the bush (vision cannot pass through opaque barriers). From Povinelli and Eddy (1996b). Copyright 1996 by Cambridge University Press. Reprinted by permission.
ideas to a test by having our chimpanzees enter a test unit, approach an experimenter, and use their species-typical begging gesture to request a food item (Figure 3.7). In the crucial test condition, as soon as the ape gestured, the experimenter leaned and looked along a predetermined line of sight that struck an opaque partition that separated the chimpanzee and the experimenter (glance to partition, Figure 3.7). The line of sight was carefully planned so that if projected through the partition, it would strike the back wall of the testing unit. If the apes were just responding to the orientation of our eyes, head, and upper torso, and, as a result, turning to scan along an imaginary trajectory, they should have ignored the interruption in the experimenter’s line of sight. They should have projected the experimenter’s

line of sight through the opaque partition and wound up looking at the back wall of the test unit. In fact, they virtually never did. In support of the predictions generated by the high-level model, in their first two encounters with these conditions, the apes attempted to look around to the experimenter's side of the partition (Povinelli & Eddy, 1996b, Experiment 2; see Figure 3.8). Several control conditions were employed. For example, in one condition, as soon as the ape gestured to the experimenter, the experimenter engaged in an irrelevant movement for 5 seconds; in another condition, the experimenter instead sat still, staring at the subject for 5 seconds before handing over the food (see Figure 3.7). In these conditions, the apes virtually never looked around the partition.

Does this mean that the high-level model is correct—that chimpanzees understand gaze as the projection of the mental state of attention? Again, maybe, but maybe not. The more we considered these results, the more we came to the conclusion that there were really two alternative explanations. One, of course, was that the high-level model was correct. The other was that the low-level model was simply too low-level. In other words, perhaps it failed to consider for how other mechanisms—most notably, general learning—might interact with the gaze-following response. If the low-level account in Figure 3.6 is accurate, then surely with enough experience following the gaze of others, chimpanzees will begin to form measurable anticipatory reactions (e.g., changes in heart rate) when they see someone look off in an unusual direction. Of course, these anticipatory responses will become paired or associated with the changes that occur during the orienting response that results when they see what the other was looking at (see Figure 3.6). In simple terms, then, regardless of what chimpanzees (or young infants) understand about attention as a mental state, when they track someone else's gaze, their neurophysiological systems certainly will anticipate receiving interesting or novel stimulus information. Thus, it does not require understanding gaze as a projection of the mental state of attention to become conditioned to look on the other side of an opaque barrier. After all, in the world of the subjects' previous experiences, this is where the interesting events or objects have always been located. With enough experience in situations of this sort, the apes will be conditioned to look on the other side of opaque obstacles. Thus, the diagnostic response on our test may be accomplished through the interaction of the orienting reflex, general learning mechanisms, and a "simple" gaze-following mechanism. Is the resulting behavior that emerges from these fairly simple mechanisms unimpressive and trivial? Absolutely not. But this should not dupe us into prematurely heralding these results as unambiguous evidence of a theory of attention in chimpanzees. Indeed, as will be evident shortly, we have additional reasons to be extremely skeptical.
FIGURE 3.8. Two views of a chimpanzee responding to gaze and body motion in the direction of an obscured target (glance-to-partition, see Figure 3.7). In (A)–(C), the experimenter looks at a target on an opaque partition so that her line of sight (if projected straight ahead) strikes the back wall of the test unit. In response, the subject leans forward and sideways, apparently attempting to look at the surface of the partition facing the experimenter. In (D)–(E), the same trial is shown but from a videotape view shown to naive raters. Raters were told to note whether the subject attempted to look at the square target. The chimpanzees demonstrated an ability to appreciate the interaction between an experimenter’s line of regard and an opaque surface (see Figure 3.7 for control treatments).
Do Chimpanzees Understand Seeing-as-Attention?

Over the past several years, our laboratory has conducted dozens of studies to examine what, if anything, chimpanzees know about seeing-as-attention. We initially began by focusing on their natural begging gesture described earlier (Povinelli & Eddy, 1996a; Figure 3.9B). The gesture is used in several communicative contexts, including situations in which one ape is seeking reassurance from another, or in cases where one ape is attempting to acquire food from another. Every day, our apes spontaneously use this gesture to request treats such as bananas, apples, sweet potatoes, onions, or carrots from their caretakers and trainers. It occurred to us that we might be able to use this gesture to investigate whether or not chimpanzees understand the attentional aspect of seeing. We first trained them to enter the test unit and look to see whether an experimenter was positioned on the right or the left (see Figure 3.9). With food just out of reach, the natural response of the apes was to look at the experimenter and beg through

FIGURE 3.9. Although young chimpanzees are very sensitive to the eyes of others, they do not appear to understand their attentional significance in the manner that 2½-year-old humans do. Chimpanzees can be trained (A and B) to use their natural begging gesture to request food from an experimenter. If an experimenter stands on the left (A and B), the chimpanzee will gesture through the hole on the left side of the partition; if the experimenter stands to the right, the subject will respond to the right. However, this is no guarantee that the subjects realize that this person is subjectively linked to them via the mental state of attention. When two experimenters are used, one who can see the chimpanzees and one who cannot (C), the chimpanzees typically respond randomly. The apes apparently fail to understand that only one of the experimenters is visually and mentally connected to them, and thus able to provide food. Perhaps this is because they simply cannot represent mental states of attention. Alternatively, chimpanzees might simply fail to understand the specific role that eyes play in deploying attention. From Povinelli and Preuss (1995). Copyright 1995 by Elsevier Trends Journals. Reprinted by permission.
the holes in the Plexiglas toward the experimenter or food. We trained the apes to be selective about where they would gesture. If the experimenter was on the right, and the ape begged through the hole on the right, the experimenter handed him or her a piece of fruit or small cookie. If the ape gestured through the wrong hole, he or she was ushered back outside and another trial started. Once the apes were excellent at this, we were in a position to begin to ask them what they understood about seeing.

To accomplish this, we occasionally confronted the apes with not one experimenter, but two—one of whom could clearly see the ape, the other of whom could not. In considering the best way to go about this, we first selected objects and toys with which our apes had extensive experience. In particular, we selected objects that we had seen them using in ways that suggested they might understand something about seeing. For example, one of our apes’ favorite games was to take large plastic buckets, pull them over their heads, and then walk around their enclosure bipedally. With one hand stretched out in front of them, and the other holding the bucket in place, they would amble about until they bumped into something. Indeed, in the context of these antics, they would even occasionally lift the bucket slightly: to peek, as it were. Another routine involved draping burlap sacks or other cloths over their heads and performing similar stunts. Some versions of these games did not rely on using objects at all. One, for instance, was quite simple: They would take the palms of their hands and cover their eyes. It is important to stress that none of these observations definitively established that our apes understood seeing (either in themselves or in others). Clearly, a sober dissection of these behaviors did not force us to accept this interpretation. But the subjective impression was unavoidable—surely, they at least understood seeing as an event that connects subject and object together via some kind of attentional glue (i.e., Level 1 perspective taking; see Figure 3.3a).

When testing began, the subjects entered the test unit and performed excellently on the trials involving a single experimenter. As they had previously learned, they approached the Plexiglas in front of the experimenter and requested a piece of food. In a preliminary control phase, the apes entered the test unit and encountered two familiar experimenters. One of them was holding out a desirable piece of food, and the other was holding out an undesirable block of wood. After taking a look, the chimpanzees responded by gesturing to the one offering the food. Thus, when the apes did not have to reason about who could see them, they had no trouble choosing between two people.

But what about the crucial trials—those on which the apes were confronted with two experimenters, one of whom could see them, the other of whom could not? Using the spontaneous games of the apes as a guide, we constructed several ways of posing this question. In one case, one experi-
menter held a bucket over her head, whereas the other experimenter held a bucket in such as way as to still allow her to clearly see the subject. In another condition, one experimenter covered her face with her hands, the other covered her ears. In still another condition, one tied a blindfold over her eyes, whereas the other tied it over her mouth. Much to our astonishment, the intuitive interpretation faltered, and the view counseled by sobriety prevailed. In all of the conditions just described, the apes acted exactly as if it did not matter that only one of the experimenters could see them (see Figure 3.9C). Despite their extensive and recent experience with the objects and scenarios used, the subjects gestured with equal frequency to the trainer who was visually connected to the scene and the one who was not (Povinelli & Eddy, 1996a, Experiment 1). Nor did their performance improve across the four trials they received in each condition. There was, however, a single, notable exception. One of the phases involved probe trials in which one trainer faced forward and the other faced backward. In direct contrast to all of the other conditions, here the apes gestured to the trainer facing forward from their very first trial onward. With this puzzle on our hands, we embarked on an additional 13 experiments to determine what factors the chimpanzees were using to make their selections.

Initially, we focused on determining why the apes had such an easy time with the back-versus-front trials, but such a difficult time with all of the others. The first idea that occurred to us was that back-versus-front trials were just a more natural or obvious instance of seeing versus not seeing. But we knew there was an alternative, more mundane explanation lurking nearby: Perhaps the apes were selecting the experimenter who was facing forward because that was precisely what we had initially trained them to do. From their perspective, in this condition one of the two people was configured in exactly the same manner as in training. In contrast, the other person was not. Thus, they may have simply been rehearsing a learned procedure to gesture to the frontal stimulus of a person. We realized that one way of subjecting these alternative explanations to a test would be to confront the apes with a new condition in which both of the experimenters faced away, but one looked back over her shoulder in the direction of the subjects (Figure 3.10). If the mentalistic interpretation were correct, the chimpanzees could be expected to gesture to the person looking over her shoulder. In contrast, if the apes’ success on the original back-versus-front condition was based on a rule about a frontal stimulus, they ought to fail to discriminate between the two. Much to our surprise, this latter outcome is exactly the result we obtained (Povinelli & Eddy, 1996a, Experiment 3).

Despite the fact that the lower-level framework had generated some correct and impressively unexpected results, we did not abandon the idea that our chimpanzees might still understand that one of the experimenters was attentionally connected to them. Indeed, we conducted 11 additional
studies in which we varied the postures, locations, and eye direction of the experimenters in a systematic effort to uncover some hidden understanding on the chimpanzees part. Across these studies, it became clear that the subjects’ performances were improving—at least in some conditions. However, further research revealed the probable reason for this improvement: The subjects were learning a rule that could be verbally summarized as, “select the person whose face is visible.” Thus, the apes began to perform well in those cases (such as the one involving buckets) where one of the experimenters face was completely absent (in this case, obscured by the bucket) and the other was clearly visible. Yet they continued to perform poorly when equal amounts of the faces were visible (e.g., blindfolds over the eyes
versus the mouth)—despite the fact that only one of the experimenters could see them (see especially, Povinelli & Eddy, 1996a, Experiment 13). Indeed, after additional experience the apes even began to learn a hierarchy of rules: “Choose the person whose face is visible, but if an equal amount of faces are visible, select the person whose eyes are visible” (see especially, Experiment 14).

At this point, one might wonder what the difference might be between these chimpanzees and young children. After all, simply because it took the apes a while to learn these rules, if they were ultimately able to use them properly, who is to say that they did not now understand seeing as attention? In fact, there were at least two variations of this possibility (Povinelli & Eddy, 1996a). First, our chimpanzees may have possessed a sensorially undifferentiated theory of attention from the outset and simply required experience to learn how to apply it in the context of these tests. Second, perhaps something about participating in these tests actually forced the subjects to construct an understanding of attention that was not originally present. But how were we to distinguish between these two possibilities on the one hand, and the altogether leaner interpretation that they had simply learned a series of procedural rules, with no concomitant appreciation of attention at all?

Although our laboratory has been systematically chipping away at this problem from several directions, perhaps some of the most striking data on this point inadvertently arose just over a year after we had completed the prior studies. As part of a somewhat unrelated study, we wanted to demonstrate that our animals would respond appropriately to an eyes-open-versus-closed condition that we had used in one of the earlier investigations. By the end of the Povinelli and Eddy (1996a) studies, our subjects had displayed some evidence of responding selectively to faces in which the eyes were visible (Experiment 14). But 13 months later, as we readministered the eyes-open-versus-closed trials to our subjects, an outcome we had not anticipated emerged—the animals performed at chance levels. Even after receiving 48 trials of this type, they continued to respond randomly. Indeed, using another, less subtle condition (screen-over-the-face) to which they had previously learned to respond excellently, and with which they had far greater experience, the subjects’ performances only gradually improved, barely bobbing above chance after 12 trials (Povinelli, 1996b). Their failure to show stable evidence of retention cast grave doubt on the idea that the chimpanzees had acquired a theory of attention during their participation in the initial round of research.

After pondering the contrast between the quite sophisticated gaze-following abilities of our subjects, and their relatively poor performance on the tests just described, we embarked on another set of studies to look for evidence that our apes might understand attention in other contexts
(Povinelli et al., in press). It occurred to us that perhaps the apes had trouble reasoning about the differing visual perspectives of two persons simultaneously, or perhaps they had difficulty understanding themselves as objects of the visual perspective of others. Given their excellent gaze-following abilities, we reasoned that they might show better evidence of understanding the “aboutness” of visual perception if we created the following situation. First, we trained the apes to search under two cups for a hidden treat. In the initial stages, the experimenter was equidistant between the two cups and pointed to one of the cups while staring at a target positioned midway between the two. After considerable trial and error, the subjects learned to use the experimenter’s pointing behavior to select the correct cup.4 This set the stage to test the chimpanzees by occasionally having them enter the test lab and discover that the experimenter was no longer pointing, but instead was gazing in one of three ways: at the correct cup (At Target), above the correct cup (Above Target), or at the correct cup but just using the eyes (Eyes Only) (see Figure 3.11). Our reasoning was simple. If the chimpanzees understood the referential significance of the gaze of the experimenter, they ought to select the correct cup on the At Target trials and possibly the Eyes Only trials, but should choose randomly between the two cups on the Above Target trials. The latter prediction is the key one, because we reasoned that organisms with a theory of seeing-as-attention (e.g., 3-year-old children) would interpret the distracted experimenter as not conveying any information about the location of the reward.

In order to evaluate our task, we first tested whether 3-year-old children met the central prediction of the mentalistic framework. As we predicted, these children selected the cup that the experimenter was looking at on the At Target trials, but chose randomly between the two cups on the Above Target trials (Povinelli et al., in press, Experiment 3). This result is crucial, because it demonstrates that our theory of the task was supported. In direct contrast, the chimpanzees did not discriminate between the At Target and Above Target trials. Rather, they entered the test unit, moved to the side of the apparatus in front of the experimenter’s face, and then chose the cup that was closest to them. Did the apes simply not notice the exact direction of the experimenter’s gaze on the Above Target trials, thereby confusing these with the At Target trials? To the contrary, the results indicated that the apes looked above and behind themselves on over 71% of all Above Target trials, a level dramatically higher than on either the At Target or the Eyes Only trials (16 and 7% of the trials, respectively; Povinelli et al., in press, Experiment 1). Thus, although the chimpanzees clearly noticed and responded to the gaze direction of the experimenter in the Above Target trials, unlike 3-year-old children, they provided us with no unique reason to believe that they understood how this posture was connected to an internal state of attention.
FIGURE 3.11. Experimental apparatus and conditions used with chimpanzees and children to assess chimpanzees' understanding of the "aboutness" property of visual perception in the context of gaze following. In the At Target condition (A), the experimenter directs her face and gaze at the correct cup. In the Eyes Only condition (B), the experimenter directs only her gaze at the correct cup. In the Above Target condition (C), the experimenter directs her face and gaze above the correct cup. See text for details of results and interpretation.

Finally, consider an additional study we conducted in which our chimpanzees were asked to reason about one person's attentional state. This study built on some previous work by Gómez (1996), who constructed a situation in which chimpanzees needed an experimenter's assistance in obtaining a food reward. Gómez varied the state of the experimenter's attention on each trial (e.g., eyes open vs. closed). He reasoned that if chimpanzees understood the differential attentional states of the experimenter in these situations, they ought to differentially deploy their spontaneous attention-getting behaviors. Indeed, Gómez reported that his 4-year-old chimpanzees did engage in attention-getting behaviors when an experi-
menter was inattentive. However, from his brief report of these findings, there is no evidence that the deployment of these behaviors was specific to these inattentive conditions, as opposed to other, attentive conditions. Thus, the data he has presented show that chimpanzees possess and deploy attention-getting behaviors. They in no way uniquely demonstrate that they understand the connection between these behaviors and the internal mental state of attention in the other person.

In order to address this question directly, we created several conditions in which we manipulated the state and location of the experimenter’s attention. On most trials, the chimpanzees simply entered the test unit, gestured to the experimenter, and were immediately handed a food reward. On the crucial trials, however, the experimenter activated a 20-second timer as soon as the subject gestured to them. During these 20 seconds, the experimenter engaged in one of several behaviors. He or she either (1) stared directly at the subject while attempting to make direct eye contact, (2) made direct eye contact with the subject while engaging in slight back and forth movements of the head (a signal of attention in chimpanzees), (3) closed their eyes and waited, or (4) looked above and behind the chimpanzee. In the first two of these conditions, the experimenter maintained a state of visual attention to the subjects throughout the 20-second waiting period, whereas in the latter two, the experimenter was visually inattentive for the waiting period. We predicted that if the subjects appreciated the different state of the experimenter’s attention in these conditions, they ought to deploy more attention-getting behaviors (touching/slapping at the experimenter or vocalizing) in the inattentive conditions than in the attentive ones. In fact, although the subjects readily engaged in such attention-getting behaviors, they did not exhibit more or longer episodes of them in the inattentive conditions as compared to the attentive ones (Povinelli, Davidson, & Theall, 1998).

Do Chimpanzees Understand Seeing-as-Knowing?
The results just described cast doubt on whether chimpanzees understand seeing-as-attention. It is curious, therefore, that earlier reports suggested that chimpanzees might understand even some of the more sophisticated aspects of visual perception. For example, can chimpanzees understand that someone who sees an event possesses knowledge different from someone who does not? Premack (1988) and Povinelli, Nelson, and Boysen (1990) both reported studies that attempted to address this question. However, both studies possessed serious methodological limitations that prevent a strong inference that chimpanzees were reasoning about knowledge states per se. In the Povinelli et al. (1990) study, chimpanzees were allowed to choose between the pointing “advice” of two experimenters, a “know-
er" who had previously seen where food was hidden, and a "guesser" who had been out of the room during the hiding procedure. Although 3 of the 4 subjects showed some evidence of understanding the task, this understanding did not emerge until after they had experienced many trials. Indeed, even when the subjects were confronted with a novel version of the task, evidence for comprehension did not emerge until after they had received a number of trials (Povinelli, 1994b). Thus, there is ample room to suppose that their performance was the result of trial-and-error learning. In contrast, and consistent with the findings of other developmental psychologists using different procedures, 4-year-old children performed excellently on this task from Trial 1 forward (Povinelli & deBlois, 1992b). (Other critiques of these studies are provided by Povinelli & Eddy, 1996a, and Tomasello, 1996.)

Other Mental States, Other Tasks, Important Caveats

Clearly, there is more to our understanding of the mind than our reasoning about the connection between visual perception on the one hand, and attention, knowledge, and belief on the other. Indeed, as we have seen, developmental psychologists have been exploring infant and young children's understanding of a wide range of mental states and activities including intentions, desires, emotions, inference, stream of consciousness, and false belief, just to name a few. Yet even though the theory of mind concept originated in the context of research with chimpanzees, current research with nonhuman primates has lagged behind. However, there have been a few attempts to investigate related topics such as their understanding of the distinction between accidental and intentional behaviors (Povinelli, Perilloux, Reaux, & Bierschwale, in press). In general, there is little evidence that nonhuman primates reason about other mental states any more than they do about attention (see Tomasello & Call, 1997, for a thorough review of the topic).

An important caveat concerning development is necessary in attempting to draw any firm conclusions about chimpanzees' understanding of mental states. In most of the studies reviewed here, and indeed in virtually all of the research to date, the age of the subjects has not been an independent variable. Thus, the majority of the tests we have described took place when our subjects were between 5 and 8 years of age. Admittedly, the abilities we were testing for are well consolidated in human children by 2 or 3 years of age. However, there is no a priori reason to suppose that even if there is extensive overlap in the cognitive developmental pathways of humans and chimpanzees, the rate of development is the same. Indeed, there are some data to suggest that certain behavioral capacities (including self-recognition in mirrors) that emerge at about 18–24 months in humans are
not acquired by chimpanzees until about 5–8 years of age (for a partial re-
view, see Povinelli, 1996a). Thus, it is still possible that adult apes 
might be reasoning about emotions, desires, and goals, for example, in a 
manner comparable to 18- to 24-month-old human infants (e.g., Kagan, 
1981; Baldwin, 1991; Baldwin & Moses, 1994; Brownell & Carriger, 
1990; Lewis et al., 1989; Zahn-Waxler, Radke-Yarrow, Wagner, & Chap-
man, 1992; Repacholi & Gopnik, 1997; Bischof-Köhler, 1988; Meltzoff, 
1995).

Summary: Chimpanzees’ Understanding of Mental States

We conclude that at present there is little compelling evidence to suggest 
that chimpanzees understand seeing-as-attention. Indeed, although we 
have not provided as thorough a review as we would like, there is little evi-
dence that chimpanzees understand anything at all about mental states. 
These statements, however, must not be interpreted to mean that chim-
panzees are open-ended, Skinnerian learning machines. Quite to the con-
trary. Chimpanzees, like other nonhuman primates, appear to be emotion-
ally grounded to the dynamic social interactions that cascade around them. 
They are intelligent, alert, and cognizing organisms, able to respond appro-
priately and dynamically to changing social circumstances. They possess 
communicative signals with high emotional valence, signals that mediate 
approaches, withdrawals, and a wide range of social behaviors (Goodall, 
1986; for experimental data, see Povinelli & Eddy, 1996c). They evince evi-
dence of the intelligent use of social information, such as following the 
gaze of conspecifics, and engage in flexible patterns of social interactions 
that allow them to manipulate and deceive each other. What they may not 
possess is an elaborated theory of these social dynamics—a coherent, orga-
nized body of knowledge that recruits concepts such as attention, desire, 
knowledge, and belief to explain the behavioral landscape that is constant-
ly unfolding around them.

THE SOCIAL INTELLIGENCE MODEL REVISITED

Thus far, we have examined several lines of evidence concerning what non-
human primates, and especially chimpanzees, know about self (as revealed 
through self-recognition studies) and other (as revealed through theory-
of-mind studies). The data we have just reviewed suggest a complicated 
mosaic of similarity and difference in the self and social intelligences of 
chimpanzees and humans. It is always possible, of course, to react to such 
findings as incoherent, and conclude that there must be something flawed
about the nature of the tests that have yielded this pattern. Elsewhere, we have assessed the extent to which such methodological objections have merit (Povinelli, 1996a). Rather than reexamining those arguments here, we note that as irritating as these findings may be to our preconceived notions of what a coherent account ought to look like, they are consistent with a growing body of data in related areas. For example, chimpanzees do not appear to possess genuine imitative skills (see Tomasello, Kruger, & Ratner, 1993). Although we recognize the immature nature of our current understanding of chimpanzee cognitive development, we nonetheless believe that the current findings demand some kind of evolutionary account. Thus, we ask whether there are any available theoretical frameworks that could explain the existence of such a radical incongruity between the similarity of human–chimpanzee behavioral patterns, on the one hand, and dissimilarity in theory-of-mind abilities on the other.

One possibility, of course, is the social intelligence model. As we have seen, this proposal points to the social arena as the sharpening stone against which self and social knowledge were honed. But a moment’s reflection will reveal that the social intelligence framework is unable to explain the emerging pattern of data on self and social intelligence. For example, it cannot explain why the kind of self-representational system that chimpanzees may possess would have evolved in the common ancestor of the great apes and humans, but not earlier (see Povinelli & Cant, 1995). There are many ways of measuring social complexity, but however we choose to measure it, great ape societies do not stand out as being exceptionally sophisticated. Indeed, by certain measures, such as group size, baboons stand out as facing more social challenges per unit time than chimpanzees (Dunbar, 1988). For example, baboons display patterns of social intrigue every bit as captivating as chimpanzees (e.g., Smuts, 1985; Strum, 1987; Whiten & Byrne, 1988). Thus, if there are phylogenetic differences in self-representational abilities among nonhuman primate species, then attempting to map them onto differences in social complexity will immediately run into trouble.

The social intelligence framework fares even less well when it comes to explaining the evolution of theory-of-mind abilities. The first difficulty is the mounting evidence (some of which we have reviewed) that not even chimpanzees possess much in the way of an ability to represent the mental states of others. If humans are largely unique with respect to reasoning about the mental states of self and other, how can sociality—widespread and ancient in the primate order—be the adequate causal variable in explaining the emergence of theory-of-mind skills? Clearly, there must be some other factors or conditions that explain the restriction of this kind of cognitive system to the human species.
THE EVOLUTION OF SELF-AWARENESS

EVOLUTION OF THE SELF–OTHER FUSION: ALTERNATIVE HYPOTHESES

We have seen that humans appear to develop an understanding of the minds of others that parallels their understanding of their own. However, the exact connection between these conceptual knowledge bases is unclear. As we have seen, the theory-theorists argue for a kind of general theoretical system that is applied uniformly to self and other in developing ideas about intentional states (e.g., Gopnik, 1993). In contrast, many simulation theorists believe that information about the self's mental states is primary, whereas knowledge of comparable states of others is derived from such knowledge. Nonetheless, both theoretical positions acknowledge the intimate connection between self understanding and an understanding of others. To date, however, no one has seriously examined the possible evolutionary history of this fusion of self–other understanding in humans. Although we fully recognize that there are numerous possibilities, we now offer two broad hypotheses that we believe raise the central questions connected with this problem.

The Asynchrony Hypothesis

In principle, it is possible that despite the intimate connections between self–other knowledge in our species, the evolutionary history of these knowledge systems was separate. Although either sequence is logically possible, we wish to explore the possibility that knowledge of the self's intentional states evolved prior to knowledge of similar states in others.

In light of the failure of the social intelligence framework to account for the phyletic differences in self-recognition, Povinelli and Cant (1995) offered a hypothesis to explain why these abilities are restricted to the great apes and humans. Although we do not have space to review this proposal in detail here, they hypothesize that a large evolutionary increase in body weight, coupled with a commitment to an arboreal lifestyle, created unusually severe locomotor problems for the Miocene ancestor of great apes and humans. They speculate that the solution to this problem was the adoption of a nonstereotyped form of locomotion called orthograde clambering. This form of locomotion may have required an explicit representation of the self's actions (i.e., an explicit understanding of the self's agency). As we described earlier, this may be exactly the kind of self-representational system that allows for self-recognition in mirrors (see Povinelli, 1995). Thus, Povinelli and Cant's "clambering hypothesis" explains, from an evolutionary perspective, why self-recognition appears limited to organisms within the great ape–human group.

Their model also raises the possibility that this primitive representa-
tional system is not uniformly applied to self and other. In others words, although the extent of self-knowledge present in chimpanzees is debatable, if something like Povinelli and Cant's (1995) model turns out to be correct, it raises the possibility that selection may have favored a qualitatively new mechanism for representing the self's actions, without a concomitant alteration in their representation of others. If true, and if great apes have not undergone extensive psychological evolution in this regard since that point, the emergence of an integrated self-other representational system may have occurred at some point during the period of rapid brain size evolution during the course of the last 2 million years of human evolution (see Preuss & Kaas, in press). If these claims turn out to be correct, then the intimate psychological relation between self and other may turn out to be one of the key psychological distinctions between humans and their closest living relatives. On this view, knowledge of the mental states of self and other evolved in asynchrony.

The Synchrony Hypothesis

Alternatives to the asynchrony hypothesis can be thought of as a family of hypotheses, which all share the common assumption that an interpretive stance toward self-other knowledge evolved in lockstep at each evolutionary juncture. For example, one possibility is that the presence of mirror self-recognition in chimpanzees indicates the presence of an explicit (albeit limited) self-concept, as well as a parallel understanding of others. Thus, if the evolution of such a self-concept evolved in the ancestor of the great apes and humans (e.g., the clambering hypothesis), this may have resulted in an immediate, correlated understanding of the actions of others. This particular hypothesis commits one to the view that chimpanzees and other great apes that display evidence of self-recognition possess an explicit self-concept. However, an alternative version of the general synchrony hypothesis is that great apes possess only implicit representations of self and other. In other words, Povinelli's (1995) model that explicit self-representations of one's actions are necessary for self-recognition in mirrors may be wrong. Instead, great apes may develop only implicit self-representations (although these representations must somehow differ from those possessed by other nonhuman primates in order to yield a system capable of engaging in self-exploratory actions in front of mirrors). Thus, in its broadest construction, the synchrony hypothesis concerns the uniformity of self-other representation, not the implicit-explicit nature of the representations.

In an era in which dedicated (if hypothetical) brain modules have somehow become a favored account of cognitive evolution, some may have trouble believing that such an understanding would evolve uniformly. However, if selection favored general representational structures, then re-
gardless of whether the local reason for these evolutionary changes had to do with the need to explicitly represent the self's actions in relation to environmental effects of the self's actions (the clambering hypothesis), the consequences would be systemwide. Although current data have failed to produce compelling evidence for theory-of-mind abilities in chimpanzees, if the synchrony hypothesis is correct, and if the common ancestor of the great apes and humans evolved a limited understanding of self and other as mental agents (perhaps somewhat akin to the ways in which 18-month-old humans understand self and other), then Gallup's (1982) proposal could still be rescued. For example, if chimpanzees could be shown to understand attention in the ways in which 18-month-old human infants appear to, it would still be possible that self-recognition in mirrors might be a relatively easy-to-detect marker of a strong inferential connection between the psychology of self and other. 5

REINTERPRETING BEHAVIOR

There remain several nagging ambiguities in our account. If, as we have suggested, chimpanzees possess a limited self-conceptual system, and at best possess only very circumscribed knowledge of other minds, then how can we account for the remarkably similar behaviors between our species? Perhaps the point is best put baldly: If we admit that our species reasons about mental states when we engage in a given behavior, how can we deny other species a similar understanding when they engage in similar behaviors? Conversely, in this sense our critique of Romanes' (1882) method can be stood on its head: If the complicated social intrigues practiced by chimpanzees and other nonhuman primates can be accounted for without granting them an understanding of mental states, why do we insist on this account of our own?

The mistake, we believe, lies in our understandable tendency to start with human behavior and psychology and work our way to other species. As an alternative, envision our planet long before humans or the modern great apes evolved. Envision a wide variety of language-less primate species negotiating their way through a complicated social milieu. Imagine untold generations of descendants evolving ever more sophisticated social rules and procedures—some tightly constrained developmentally, others more open-ended. And let us be especially careful not to oversimplify those abilities. Grant these species rich and diversified behavioral systems—systems composed of many behaviors still present in us. But now let us imagine that these species were devoid of any understanding of themselves or others as mental agents. Clever brains, in Humphrey's (1982) turn of phrase,
but blank minds. Next, imagine the first spark of self-awareness emerging in common ancestor of the great apes and humans—perhaps a limited understanding of the self as causal agent (Povinelli & Cant, 1995). Finally, imagine that only one of that ancestor’s living descendants—the genus Homo—evolved an additional cognitive specialization that produced a uniform, mentalistic understanding of self and other. What we are suggesting is that at some point during the emergence of hominids as a distinct group of species, a new conceptual understanding of others may have been woven into existing, ancestral developmental pathways controlling the expression of myrad ancient (but sophisticated) social behaviors.

In short, our hypothesis is that the terminal addition model of general psychological evolution proposed by Parker and Gibson (1979) cannot be fruitfully applied in the domain of psychological systems that we call theory of mind. Consider the self-other psychological systems that we are suggesting could be unique to humans. Rather than having tacked these systems onto the end of the cognitive developmental pathways that were present in the last common ancestor of humans and chimpanzees, we speculate that these skills were woven into those pathways at some point much earlier in development. If correct, human psychological evolution could be characterized less by the addition of new behaviors than by the weaving in of new cognitive systems alongside (and probably into) the old. Like a tapestry into which new colors were added, the old tapestry was not discarded or merely lengthened. Rather, the ancient neural systems may have served as both a substrate and/or constraint for the new systems. And, as we elaborate later, these new systems, in turn, may have had the effect of functionally reorganizing the older ones.

At this juncture, it may be instructive to consider a similar problem faced by researchers interested in cognitive development. During the course of development, young children often display the following enigmatic behavior. On the one hand, they will perform a given task immediately, effortlessly, and correctly; on the other hand, when queried about their actions, they will often offer a lucid, quite precise, but utterly irrelevant explanation for these actions. Indeed, we have discussed one example of this phenomenon in the context of children using obvious information to locate hidden objects, but failing to understand how this information assisted them in the first place. These kinds of mismatches between production and comprehension have led Annette Karmiloff-Smith (1992) to propose a process of “representational redescription” as a general feature of cognitive development. As she explains:

[Representational redescription] involves a cyclical process by which information already present in the organism’s independent functioning, special purpose representations, is made progressively available, via re-
THE EVOLUTION OF SELF-AWARENESS

descriptive processes, to other parts of the cognitive system. In other words, representational redescription is a process by which implicit information in the mind subsequently becomes explicit information to the mind, first within a domain and then sometimes across domains. (pp. 17–18)

In this way, cognitive development can be thought of as the process of recoding information that is stored in one representational format or code into a different one. Thus, a spatial representation might be recoded into a linguistic format, or a proprioceptive representation into spatial format. Each redescription, or re-representation, is a more condensed or compressed version of the previous level. . . . [the representational redescription] model postulates at least four hierarchically organized levels at which the process of representational redescription occurs. (p. 23)

It is significant for our framework that the first two of these postulated levels are not necessarily available to conscious access. Thus, Karmiloff-Smith’s (1992) account offers a particularly thoughtful explanation of how perfectly articulate children can possess and process knowledge that they cannot express—in other words, how information can be present in a system but reside at different levels of explicitness. It is our contention that the information controlling many of the common behaviors of humans and other primates inhabits these less explicit representational levels. To use Karmiloff-Smith’s terminology, one interpretation of our hypothesis is that humans have uniquely evolved mechanisms allowing for the very process of representational redescription—the process of transforming implicit representations into forms available to consciousness. 6

The significance of this view is that it dramatically alters how one interprets the behavioral and cognitive similarities and differences among humans, chimpanzees, and other nonhuman primates (Povinelli, Zebouni, & Prince, 1996; Povinelli, 1996b). Rather than interpreting such behavioral similarities as prima facie evidence in favor of the view that chimpanzees must possess the same psychological systems that attend our execution of these behaviors, it is possible to consider an alternative. Many behaviors that our species naively interprets through a mentalistic framework may have evolved and been in full operation millions of years before we appeared on the scene. We are not suggesting an extreme form of dualism (such as epiphenomenalism) in which these metacognitive states attend, but play no causative role in behavior. No, we suppose that representations of the mental states of self and other evolved because of their useful, causal connection to the behavior of the organisms possessing them. But until now, we may have been thinking in the wrong way about how the evolution of a novel psychological system may have affected the behavioral fab-
ric of human ancestors. To date, many researchers (including ourselves) have been looking—largely unsuccessfully—for a simple connection between the emergence of a novel psychological system related to theory of mind and a class of novel behaviors onto which this novel system could be mapped. But this may turn out to have been a fool's quest. The evolution of a new psychological system may have functioned to support, optimize, or otherwise reorganize existing behavioral patterns, without necessarily leading to any individual elements that look definitively novel. Thus, although it may have had a profound effect on the complexity, rapidity, and informational density that a behavioral network could achieve, it may not have sprouted many "new" or "novel" behaviors per se. We shall return to this point later.

**Lessons from Gaze Following**

In order to illustrate our argument, we return to the case of gaze following—although we intend to show how the same logic may apply with equal force to other behavioral systems as well. First, when we follow the gaze of someone else, it is not clear when, exactly, we begin to entertain notions about the other person's internal mental state. Indeed, there are surely many cases in which we respond by looking where they are looking and then back again, without ever even registering the fact that we have done so. Thus, in principle, it is possible that the attribution of attention to the other person might occur after the behavioral act of gaze following, thus eliminating it as a possible causal explanation of the behavior itself. Some may object here, noting that the case of adults may have little to do with infants. For example, adults may have routinized the gaze-following response after considerable experience. But this already grants a large portion of our argument by demonstrating that the two elements—the response and the attribution—are, at least in principle, dissociable. In other words, imagine that we could selectively cripple the ability to represent the mental state of another's attention. Would the gaze-following response still occur? We suspect so.

Our suggestion is that gaze following evolved not as part of a theory-of-mind system, but because of its utility in the social ecology of our primate ancestors. For example, Povinelli and Eddy (1996a, 1996b) have proposed that gaze following may be a fairly ancient behavioral mechanism common to many social primates (and perhaps other social mammals) and may have evolved independently of an understanding of the mental state of attention. A psychological system that could rapidly process information about other group members' line of sight could be very advantageous in several naturalistic contexts such as the early detection of predators and anticipating the likely targets of the social behaviors of other group mem-
bers (see Chance, 1967; van Schaik, van Noordwijk, Warsono, & Sutriono, 1983). To be sure, this system might mediate the perceptual input (the visual signal of seeing a group member turn and look) and the motor output (turning and looking in that same direction) with representations of the other's mental state (i.e., attention). However, there are other, more direct linkages that might mediate the input–output relation in question—ones that do not depend on representations of mental states per se. Thus, we propose that it is quite possible that in the context of the evolution of a broader psychological system, metacognitive representational systems were stitched into the broader psychological system of the 18-month-old, a system that already possessed the neural mechanisms supporting gaze following (see Figure 3.12). Thus, it may be accurate to say that the infant's representation of the other person's attention accompanies the act of gaze following; but we offer the hypothesis that this representation is not causally related to the triggering of the action.

At this point it may seem as if we are dangerously close to invoking a spectator-consciousness effect: The presence of an awareness of a certain action without this awareness having any causal effect in the real world. Quite to the contrary, we suppose that infants' growing abilities to represent and reason about attention (and other mental states) in those around them offers them a coherent explanatory account of what is occurring (see Gopnik & Meltzoff, 1997). This explanation in turn allows them to learn about relations among people, objects, and events in a faster and more direct manner. In the context of gaze following, we imagine that during its earliest evolutionary manifestation, the response may have functioned as a fairly automatic response to sudden shifts in the head movements of other conspecifics and may have then operated as described in Figure 3.6A. However, a first evolutionary step toward understanding gaze as attention may have involved understanding shifts in others' gaze as behaviors that signal important impending relations between the organism and other events. Indeed, it is quite possible that even chimpanzees acquire this kind of understanding of gaze, as opposed to a mentalistic understanding of gaze as a psychological spotlight emanating from the other organism.

A glaring weakness in our account is that, at present, we cannot precisely specify the nature or scope of the neural systems mediating the gaze-following response, or the attribution of attention. For example, until now, we have used metaphors related to sewing and weaving to help us convey the idea of the interleaving of new systems alongside older ones. Yet this metaphor fails in at least one crucial respect. It suggests that the newer structures could be easily pulled out, leaving the old ones intact. However, the emergence of new representational systems may often involve the direct use of ancestral neural architecture, thereby resulting in a functional inter-dependency between the old and the new. Some cases of human psy-
chopathology may help to illustrate this point. Autism, for example, has received considerable attention recently as being characterized by a severe deficit in theory-of-mind and other metacognitive abilities (e.g., Baron-Cohen, 1995). Autistic individuals have been shown to be profoundly impaired on tasks relating to an understanding of mental states such as desires, knowledge and beliefs. Interestingly, there is evidence that the inability of an 18-month-old infant to follow an adult’s gaze (accompanied by a pointing gesture) predicts a diagnosis of autism at 3 years of age (Baron-Cohen et al., 1996). Given that the dysfunction of gaze following in autis-
tic infants and children predicts a later dysfunction in their understanding of attention, one interpretation of this finding is that gaze following is, in fact, causally related to an understanding of the mental state of attention.\textsuperscript{7}

However, this conclusion need not follow. It is possible that during the course of recent human psychological evolution, new and old structures have combined into related functional systems in such a manner that it is not easily possible to selectively impair the new elements in this system. Some researchers, for example, have recently shown how the suite of behavioral and cognitive impairments associated with autism may in fact be the result of damage very early in development—perhaps as early as the fourth week of gestation (Rodier, Ingram, Tisdale, Nelson, & Romano, 1996). Thus, many of the cognitive dysfunctions of autism may be the result of abnormal inputs from the brain stem to a relatively normal forebrain. Conversely, it is possible that there are some as-of-yet undetected abnormalities in the forebrain that are secondary, cascading consequences of such early injury. In either case, the implication is clear: The evolution of new psychological systems may neither replace nor sit insulated alongside ancestral systems. Rather, new systems or subsystems may be created by building inside ancestral templates in such a manner that most of the useful behavioral propensities of ancestral organisms are conserved.

**Broadening the Explanatory Framework**

We suspect that a wide class of social behaviors can be usefully thought of in this same manner: deception, reconciliation, selective retaliation against less dominant allies of one’s aggressors, certain forms of social learning—indeed, most of the core elements of the fabric of human (and ape) behavior. Consider Frans de Waal’s (1982, 1996) marvelous expositions of the striking commonalities in chimpanzee and human politics and morality. He interprets these behavioral similarities as indicating not just basic psychological similarity between us and them, but also similarity in the kind of metacognitive processes that attend those behaviors in our own species. In contrast, we are suggesting that there are both real, and “really real” explanations for our behaviors. Certainly our folk psychological explanations count for something; indeed, they may reflect the discovery of an important shorthand for reasoning about enormously complicated neural processes. Rather than having to specify that so-and-so has such-and-such a configuration of neural activation in the prefrontal cortex, we can simply say, “She thinks that I want to trick her.” But although accurate in a sense, it may not be at this level of explanation that the utility of these representations is to be found. Indeed, many (if not most) of the behavioral propensities now associated with our species’ higher-level metacognitive representations were present long before these representations were even possible.
But if the view that we are exploring here is correct, then the striking *behavioral* similarities between humans and chimpanzees, for example, are only a guide to superficial psychological similarities. Similarity at the level of interpretation and meaning—folk psychology—is another matter altogether.

There is another manner in which our folk narratives constitute a real explanation for why we behave in the manner we do. To the extent that we form beliefs about the causative role of mental states in behavior (i.e., to the extent that we ponder why so-and-so has done such-and-such), we quickly construct elaborate narrative accounts of why events have taken the shape they have. These narrative accounts may serve as a database in their own right, a storehouse of information not so much different from information received and translated through the primary senses. Thus, in considering what we may do next, we appear to take into account not just what has happened, but why (at least from the perspective of our narrative) it has happened. Such narratives, while not strictly "accurate," may provide human beings with a powerful adaptive device, a means of rapidly reorganizing existing fundamental behavioral units into novel cultural configurations. In this sense, the diversity of human cultural beliefs and nonmaterial traditions may be the construction of narratives well suited for the current and historical ecological challenges faced by individual cultures. If folk narratives function in this fashion, this may help to explain why human populations have expanded into virtually every conceivable ecogeographic zone on the planet, whereas chimpanzees and other great apes have remained restricted to the tropics and neotropics, and even there inhabit very few niches (tropical rain forest, open woodland savannas). Indeed, part of the consequence (and perhaps function) of such causal narratives is that they open up the possibility of mental time travel, the ability to understand how the past, present, and future are all part of a flow of connected events (see Suddendorf, 1994; Povinelli, 1995). Nonetheless, despite the diversity of human habitat and culture, and despite the enormity of possibilities opened up by these cognitive specializations, we suspect that with but rare exceptions the fundamental behavioral building blocks that humans use to generate novel future actions have remained largely unaltered from those present millions of years ago in the common ancestor of great apes and humans (and, in many cases, even earlier).

Our approach can therefore be interpreted as an explicit alternative to Romanes' (1882) hope of using the spontaneous behavior of animals as portals into their minds. Although this approach may have fit the bill charged by the complete absence of information, it now seems doomed to miss the historical complexities of the evolutionary process. Just as the morphological structures and systems of modern organisms reflect the baggage of developmental constraints laid down in the Cambrian, so too do
psychological structures carry the ancient alongside the new. The error in Romanes' reasoning was that he assumed that the behavioral similarities between humans and other animals must reflect underlying psychological similarities. First, he assumed that an accurate introspective assessment of the mental states that accompany our own behavior could yield an accurate inference as to their cause. Worse still, he assumed it could yield an equally accurate inference for species other than our own. And although it may be true that humans form internal representations of mental states such as desires, knowledge, and belief that provide us with a useful means of anticipating what others will do, the basic blueprint of the behaviors we can respond with evolved long before those meta-level representations were possible.

As a final point, it is worth considering how this view articulates with the problem of human uniqueness. Faced with the apparent distinctiveness of the human species, philosophers, anthropologists and psychologists have offered numerous honorary titles for our species, each emphasizing some would-be unique characteristic. But from tool-making, to culture, to language, the animal kingdom has relentlessly refused to unambiguously concede any of these traits as exclusively human. The view we have offered here suggests a way out of this impasse by recognizing that we may never locate any truly unique and universal human behavioral traits. Even with respect to those phenomena where we seem to differ most from other species—cultural learning, pedagogy, ethics, and language (Premack, 1984; Tomasello et al., 1993; Povinelli & Godfrey, 1993)—the individual elements of the behaviors involved may have similar ancient precursors, revealed by observing a diverse array of other species. And in each case, the similarities may be great enough that there will always be adequate grist for the mill for those who wish forever to deny the existence of qualitative differences between humans and other species. Yet the real differences between humans and even our nearest living primate relatives may be much deeper, rooted in our interpretation of behavior. Human uniqueness may stem first and foremost from underlying cognitive—not behavioral—specializations. These cognitive differences, in turn, may then translate into a radical reorganization and redeployment of existing ancestral behavioral abilities. And it is this generativity, manifested through ancient behavioral patterns, that at a proximate level has yielded the remarkable diversity and flexibility of human culture.

EPILOGUE: AN APE'S EYE VIEW OF ROMANES' PROBLEM

If the alternative view we have sketched here is correct, we might do well to close by imagining the reaction of our apes to the questions we continue
to pose to them: "Yes," they reply obliquely through their responses to our still-clumsy tasks, "we share with you an enormous range of useful, complicated, and sophisticated social behaviors. Yes, we share with you a psychological system that is able to knit these behaviors into novel and productive strategies that serve to fulfill our goals and desires. And it is true that our emotions, mannerisms, and reactions are much like your own. We even possess a self-concept that offers us an objective perspective on our own behavior. But whatever gave you the idea that we have a theory of mind? Why do you want to believe so desperately that we are able to construct a self-other narrative like you? After all, it was your lineage, not ours, that tripled the size of its brain during the past 5 million years. It was your species, not ours, that constructed the idea that there are unobservable mental states that mediate behavior. And thus it is you," they conclude, "not us, who are in the position of reinterpreting ancient behavioral patterns in terms of mentalistic notions—notions that never even occurred to us."

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant No. RR-03583-05 to the University of Southwestern Louisiana New Iberia Research Center and National Science Foundation Young Investigator Award SBR-8458111 to Daniel J. Povinelli. Photographs and original drawings are by Donna T. Bierschwale. We thank Todd M. Preuss for offering valuable comments on the manuscript, and Karen Wright for thought-provoking discussions about these issues.

NOTES

1. This assumption may seem naive to some cultural anthropologists in that it appears to ignore a wealth of data on the unique ways in which concepts such as self are constructed across different societies (e.g., Geertz, 1973; La Fontaine, 1984; Lienhardt, 1984; Duranti, 1988). However, we follow those who recognize universal aspects of human social understanding, while simultaneously exploring the ways in which this understanding is shaped by cultural frames of reference (Hallowell, 1971; White, 1980; Heelas, 1981; Lock, 1981). Recently, there have been several direct examinations of cross-cultural understanding of mental states such as false belief (Avis & Harris, 1991; Vinden, 1996; see Lillard, 1998, for a comprehensive review).

2. A number of careful studies have failed to demonstrate self-recognition in gorillas (Lethmate & Dürcker, 1973; Suarez & Gallup, 1981; Ledbetter & Basen, 1982). These results stand in contrast to convincing evidence for this ability in a female gorilla who was reared by Francine Patterson as part of an intensive attempt to teach her sign language (Patterson & Cohn, 1994). Povinelli (1994a) has re-
viewed evidence suggesting that, compared to the other great apes, gorillas may have undergone a secondary reversal in key aspects of their schedule of physical maturation. These heterochronic processes may have had cascading consequences for aspects of general cognitive development, including their capacity for self-recognition in mirrors. However, secondary losses of derived traits generally are not the result of the loss of specific genetic instructions, but rather result from the shut-down of their expression due to changes in the rate and timing of other developmental events (Frazetta, 1975; Gould, 1977; Albrech, Gould, Oster, & Wake, 1979; for examples see Hampé, 1960; Kollar & Fisher, 1980). One consequence of this is that ancestral characteristics—traits that have been "lost" for tens of millions of years—can be reinstated by slight, abnormal changes in the developmental pathways involved. Given that gorillas presumably descended from a common ancestor of the great apes and human beings that was capable of self-recognition (Povinelli, 1987), it seems quite possible that the genetic instructions for the construction of the neural structures necessary for self-recognition still exist in modern gorillas, and can be reinstated through abnormal rearing environments such as that received by Patterson's gorilla (see Povinelli, 1994a).

3. Montgomery, Moran, and Bach (in press) have challenged the depth of 3- and even 4-year-old's understanding of gaze direction as attention. They have provided evidence suggesting that when the orientation of someone's body contrasts with the direction of their gaze, preschoolers are generally unclear as to the person's goal.

4. This is not to say that the chimpanzees necessarily grasped the referential significance of the pointing gesture. Indeed, in a separate series of studies, we explored what our chimpanzees really understood about this gesture (Povinelli, Reaux, Bierschwale, Allain, & Simon, 1997). These results indicated that the subjects used the gesture as a landmark cue, adopting the following rule: "Pick the cup that is closest to the experimenter's hand/finger." In contrast, even the youngest children we tested (26-month-olds) had no difficulty using the gesture in a referential manner.

5. As an example, consider the pointing gesture. In human infants, understanding of the attentional significance (the referential aspect) of the pointing gesture is consolidated between 12 and 15 months (see Murphy & Messer, 1977; Lemper, 1979; Butterworth & Grover, 1988; Morissette, Ricard, & Décarie, 1995). Thus, this would seem to qualify as an excellent arena to search for similarities. However, current experimental evidence suggests that although chimpanzees can easily be taught to learn to respond appropriately to pointing, they appear to interpret it in a different manner from humans. Unlike even very young human children who interpret pointing as indicating the attentional focus of the person making the gesture, chimpanzees appear to interpret the gesture in a nonmentalistic manner (see note 4). In addition, a majority of children do not display self-recognition in mirrors until 18 months (see earlier references).

6. Indeed, Karmiloff-Smith (1992) speculates that the "pervasiveness of representational redescription in human cognition is, I maintain, what makes human cognition specifically human" (p. 192). She further speculates that if the process of representational redescription is available to other species (such as chimpanzees), "the higher-level codes into which representations are translated during redescrip-
tion are very impoverished" (p. 192). As a related aside, if our account of the nature of the self-representation underwriting the chimpanzee's capacity for self-recognition in mirrors is correct, it would suggest that in this domain, chimpanzees may at least translate implicit representations of their own proprioceptive states into representations at the first level of explicitness (or E1) in Karmiloff-Smith's (1992) current model—a translation that does not occur in species outside the great ape/human group.

7. The situation cannot be quite so simple, however. Recent studies have shown that although gaze following is severely impaired in autistic people, their ability to engage in simple Level 1 perspective-taking tasks (e.g., "What toy am I looking at?") is not (Hobson, 1984; Dawson & Fernald, 1987; Baron-Cohen, 1989; Tan & Harris, 1991; Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1995). One interpretation of these findings is that Level 1 perspective-taking tasks can be solved geometrically, without understanding attention (Hobson, 1980, 1982; Baron-Cohen, 1994; Leekam et al., 1995). However, another possibility is that autistic individuals do retain an understanding of attention, but other systems that automatically process information about the eyes and head direction are crippled or impaired. In any event, if our research with apes is replicated, the opposite situation would seem to be possible as well: Organisms may possess systems for automatically processing information about gaze direction without an understanding of the mentalistic aspect of gaze.

REFERENCES


When Self Met Other


Perrert, D., Harries, M., Mistlin, A., Hietanen, J., Benson, P., Bevan, R., Thomas, S., Oram, M., Ortega, J., & Brierly, K. (1990). Social signals analyzed at the single cell level: Someone is looking at me, something touched me,


When Self Met Other


Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). Child Development, 61, 973–982.


When Self Met Other


