MORAL MINDS

How Nature Designed

Our Universal Sense

of Right and Wrong

Marc D. Hauser

6

ROOTS OF RIGHT

*But it is curious to reflect that a thoughtful drone . . . with
a turn for ethical philosophy, must needs profess himself
an intuitive moralist of the purest water. He would point
out, with perfect justice, that the devotion of the workers
to a life of ceaseless toil for a mere subsistence wage, cannot
be accounted for either by enlightened selfishness, or by any
other sort of utilitarian motives, since these bees begin to
work, without experience or reflection, as they emerge from
the cell in which they are hitched.*

—THOMAS HUXLEY

CONSIDER THE FOLLOWING TWIST on the classic trolley problems from chapter 3:

John is on a footbridge over the trolley tracks and can see that the trolley approaching the bridge is out of control. On the track, there are five chimpanzees, and the banks are so steep that they will not be able to get off the track in time. John knows that the only way to stop an out-of-control trolley is to drop a very heavy weight into its path. But the only available, sufficiently heavy weight is a large chimpanzee, sitting on the footbridge. John can shove this chimpanzee onto the track in the path of the trolley, killing him; or he can refrain from doing this, letting the five chimpanzees die (see figure on p. 308).

My own intuition—not one that I comfortably accept—is that it is permissible to push the large chimpanzee, even though in the parallel human case, it is not permissible—or, at least, less permissible—for Frank to push the large person off the footbridge. American college students share this intuition. What is the explanation or justification for the difference between humans and chimpanzees? Why does the utilitarian outcome rule for animals but not for humans? Logically, if it is impermissible to use one

Is it morally permissible for John to push the large chimpanzee onto the tracks?

life as a means to saving many, this principle should apply with equal force to human adults, infants, brain-damaged patients, and animals. Although people contrasting these cases rarely come up with coherent explanations, many allude to distinctive differences between human and animal life, including our responsibilities to members of our own and another species. These types of explanation zero in on some of the central issues surrounding current debates over animal rights and welfare. When we make decisions about the treatment of animals, we often appeal to perceived differences between our mental wherewithal and theirs. We draw a line that sets us apart from them in terms of distinctive abilities, including language, consciousness, emotion, and a sense of the future. Mark Twain held this view, and believed it raised an important conclusion about our own moral faculty: "Whenever I look at the other animals and realize that whatever they do is blameless and they can't do wrong, I envy them the dignity of their estate, its purity and its loftiness, and recognize that the Moral Sense is a thoroughly disastrous thing."[1]

Critics of the drawing-the-line view respond with cases involving our own species, pointing out that even though a newborn baby is not as conscious as an adult chimpanzee, and is as linguistically challenged as its furry cousin, few would use a newborn baby to save five others. Pointing to psychological differences between us and them doesn't work. Perhaps the difference stems instead from our emotional attachment, built over millions of years, designed to guarantee the welfare of humans but not other species. When faced with the trolley case, our emotional attachment to humans is greater than our attachment to animals, and thus our judgments shift. If this view is correct, it brings us back to the role of the Humean creature in guiding our judgments. We might imagine, for example, that the weaker our connection with the particular target animal, the stronger our conviction that we can use one life to save many; we might even shift from a permissible judgment to an obligatory one, especially if the animals are endangered. Those who see all of life as sacred never draw the line, and thus hold to the logically defendable position that if it isn't permissible to push one human to save many, then it isn't permissible to push one caterpillar, canary, or chimpanzee to save many. Those who see differences between species draw the line, and allow the utilitarian principle to rule.

Discussions of animal welfare and rights are largely orthogonal to the central concerns of this final part, but they tie in nicely to this center: What is the mental wherewithal of animals such that it informs our interest in the evolution of the moral faculty? Here's Darwin's answer: "Any animal whatever, endowed with well-marked social instincts, the parental and filial affections being here included, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well developed, or nearly as well developed, as in man." Darwin correctly suggests that animals with social instincts are the right sorts of animals for thinking about the origins and evolution of a moral sense. He is also correct in thinking that along the evolutionary path to our moral sense, nature must have added some extra accessories to the core, allowing individuals not only to care for others but to know why caring is the right thing to do, while harming is often the wrong thing to do. Darwin never provided a detailed depiction of what evolution added, nor why natural selection might have favored these extras. He did, however, leave open the possibility that minds "nearly as well developed" as ours could, in principle, have a moral sense, perhaps only in a rudimentary form, with selection acting to favor particularly moral groups. Jean-Jacques Rousseau was more explicit about the comparative issue, pinpointing a key difference between us and them, a uniquely human attribute:

"Every animal has ideas, since it has senses; it even combines those ideas in a certain degree; and it is only in degree that man differs, in this

respect, from the brute . . . It is not, therefore, so much the understanding that constitutes the specific difference between the man and the brute, as the human quality of free agency. Nature lays her commands on every animal, and the brute obeys her voice. Man receives the same impulsion, but at the same time knows himself at liberty to acquiesce or resist."[2]

To Rousseau, humans have free will, animals don't. For Thomas Henry Huxley, Darwin's henchman, many of our good and evil attributes were gifts of evolution, but our capacity to squelch the bad and promote the good through a system of ethics was largely a human creation: "Laws and moral precepts are directed to the end of curbing the cosmic process and reminding the individual of his duty to the community, to the protection and influence of which he owes, if not existence itself, at least the life of something better than a brutal savage."[3]

With this comment and perspective, Huxley parted company with Darwin, suggesting that evolutionary theory, and the comparative method more specifically, would provide few insights into our moral psychology. As it turns out, Darwin was right, Huxley wrong, and, unfortunately, many evolutionary biologists have followed Huxley's authoritative voice.[4] Shifting from phylogenetic or historical concerns to adaptive function, Darwin first imagined a reproductive competition among individuals within a group that included "sympathetic and benevolent parents" on the one hand and "selfish and treacherous parents" on the other. Realizing that the brave men who risked their lives would perish, as opposed to the selfish cowards who stayed home, he concluded that natural selection would not increase the numbers of the virtuous.[5] In contrast, stepping from within group competition to between group competition painted a different picture: "A tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good, would be victorious over most other tribes. . . . At all times throughout the world tribes have supplanted other tribes; and as morality is one important element in their success, the standard of morality and the number of well-endowed men will thus everywhere tend to rise and increase."

Darwin makes the assumption here that some sense of good will prevail over evil, creating a source of moral growth. He assumes that when

one group supplants another, the group with the higher moral calling wins. But, as history reveals, Darwin's assumption is false, unless one is willing to grant a higher moral ground to Genghis Khan, Pol Pot, Adolf Hitler, Idi Amin, Efrain Montt, and Ratko Mladic—all leaders responsible for massive genocides guided by the call of ethnic cleansing. There is, however, one sense in which Darwin was correct. If we look to the positive effects of organizations such as the United Nations, we see the spread of particularly virtuous moral attitudes, including global decreases in slavery, subordination of minority groups, child abuse, capital punishment, and the harmful treatment of animals. It is thus possible for some groups to facilitate the spread of what many consider universal rights.[6]

We can answer problems of adaptive function in at least two ways. The classic approach is to document how specific behaviors contribute to an individual's survival and reproduction. Consider, once again, the problem of altruistic behavior. For Darwin, being nice to someone else at a personal cost made little sense in light of the logic of natural selection. Not only are Mother Teresa and Mahatma Gandhi aberrations, but so, too, are those who leave tips at restaurants, take care of other individuals' offspring, and contribute to charities. These acts reduce each person's potential for self-promotion. If the Darwinian theory is right, selection should wipe out those silly enough to lower their reproductive worth— and ultimate genetic posterity—by investing in others. As the evolutionary biologists William Hamilton, George Williams, and Robert Trivers argued, we resolve this paradox if we think about behavior at the level of the gene. What appears to be genuinely altruistic, and good for the group, is actually the covert operation of selfish genes. We act nicely to kin because our genetic posterity is wrapped up in theirs. What is good for them is good for our genes. When we lack kinship, we act nicely if we have some guarantee of a reciprocated return. This isn't an act of kindness. Reciprocity is an act of self-interest, because it is driven by the expectation of a fair return: food for food, backrub for backrub, babysitting for babysitting.

From the gene's-eye view, the way to think about the evolution of moral behavior is to think selfishly. Instead of asking "How can I help you?" ask "How can my helping you help me?" In the simplest case, you would compare two strategies, moral versus immoral, and tot up the

number of babies for each. If the moral strategy wins, both in terms of reproductive output and in terms of its immunity to immoral invaders, then selection will favor the moralist and eliminate the immoralist. Life isn't that simple, but the logic of the argument is.

The second approach is to look at the source of an object's design features. Calling upon Reverend Paley's *Natural Theology*, Richard Dawkins argued that chance alone can explain neither the complicated and functionally precise design of a watch nor that of a living creature. While Paley appealed to God to account for complexity, Dawkins appealed to Darwin. While God has a vision, natural selection is blind. Natural selection builds organisms with complex design features based on a nonrandom but directionless process. Poorly designed variants are eliminated, well-designed ones favored. When we see an organism or organ with complex design, we see the handiwork of natural selection, a tinkerer that fine-tunes the raw material to the adaptive problem at hand. This argument applies with equal force to an animal's overall body shape as well as to its eyes, brain, and mind.

In the last chapter, I discussed Cosmides and Tooby's use of the design stance, to argue for an evolved cheater detector. As they suggest, a central problem for our Pleio-Pleistocene ancestors was to cooperate in the service of social exchange. When individuals engage in this kind of exchange, they implicitly or explicitly set up a social contract. Given that individuals can break social contracts by taking the benefit without paying the cost, selection will favor those who can detect such cheaters. Reverse-engineering the problem, we should find the psychological machinery required to identify cases of cheating, and Cosmides and Tooby's work suggests that we are so endowed. It is this kind of evidence that fits the logic of the adaptive design stance.

The controversy surrounding work on cheater detection ties into the second half of the "Why did it evolve?" question. By assuming that the cheater-detection system evolved as an adaptation for life among Pleio-Pleistocene hunter-gatherers, Cosmides and Tooby imply that this is a uniquely human adaptation. This is certainly one plausible story, but in the absence of observations of other animals, it remains uncontested. Proclamations about human uniqueness lie within the realm of speculation without

studies of other animals. And, as I discuss later, there are numerous cases of deception in animals, and several cases where cheaters are detected.

In this chapter, I examine which parts of the moral faculty, if any, evolved prior to the emergence of our own species. I use reciprocity as a centerpiece for discussion, both because of its prominence in our own species and because it naturally raises questions about the psychological prerequisites. To initiate and maintain a reciprocally stable relationship, individuals must recognize each other, recall what was given to whom, how much, when, and with what costs. Individuals must also recognize whether the resource was given intentionally or as an accidental by-product of an otherwise selfish goal, and whether the exchange of resources was done contingently. Like other social interactions, this form of cooperation relies upon many other abilities, including the establishment of expectations, emotional responses to actions that satisfy or violate these expectations, the capacity to acquire, follow, and enforce rules, and a sense of responsibility for the health of the relationship. And in humans, at least, these social relations often depend upon the development of a rich sense of self, empathic concern for others, and the ability to generate predictions about others' states of mind without any direct experience of their behavior. When we generate moral judgments about another's action, we make use of many of these capacities, even though we are often unaware of the underlying process. Perhaps these are the bits of psychology that Darwin had in mind when he considered the development of our intellectual powers. Perhaps these are the bits that, if nearly as well developed as in our own species, would give some animals a primitive moral sense, a capacity that we might feel satisfied in calling an evolutionary precursor.

DARWINIAN NODES OF ACTION

When the nineteenth-century physiologist Ivan Pavlov taught his dogs an association between hearing a bell and receiving food, he taught them to expect food once they heard the bell. What we don't know, however, is what exactly these dogs expected, because Pavlov never explored whether

they would have been satisfied by the appearance of any old food or whether they expected a particular kind of dog chow, and thus felt ripped off—cheated—when it was some other kind. The question then is what, specifically, do animals expect and think about prior to the occurrence of a predicted event? Whatever answer we give will not directly resolve questions of moral significance. However, because expectations are formed in the social sphere as well, it is important to understand the nature of expectation more broadly. If animals set up expectations and detect violations, then they should be capable of judging when an individual does something socially right or wrong, and, possibly, morally right or wrong.

In the 1920s, the psychologist Eduard Tinkelpaugh set out to determine whether rhesus macaques and chimpanzees create specific expectations about the kind of food shown and then hidden within a container. In one set of experiments, he concealed different kinds of food in one of two containers while a subject watched. He then placed a screen in front of the containers, hiding both from the subject's view, waited a short period of time, and then removed the screen so that the subject could search for the hidden food. Sometimes the content matched what was concealed and sometimes it did not. If they saw Tinkelpaugh hide a banana and then found the banana, they cooed with delight. If they saw Tinkelpaugh hide a banana and then found lettuce instead, they were either furious or puzzled.

We don't fully understand what it is like to have a primate experience of expectation satisfaction and dissatisfaction. But Tinkelpaugh's experiments have been repeated several times, including studies that reveal the neural code underlying matching expectation and detecting an error.[7] They show, without doubt, that the primate brain has evolved to set up expectations, anticipating outcomes that matter in terms of survival.

Here, I return to the theme set out in chapter 4 for human infants, and ask whether nonhuman animals set up expectations about actions and events, using the causes and consequences to detect violations. With apologies to Jim Watson and Francis Crick, I refer to these primitive detectors as DNA, for Darwinian Nodes of Action.

The first and most basic principle of action focuses on the capacity for

self-propelled motion. This is a starting point for discriminating between animate and inanimate objects:

PRINCIPLE 1: *If an object moves on its own, it is an animal or part of one.*

In the natural world, objects that move on their own are animals, and those that can't are either dead animals, plants, or inanimate objects. When animals see these kinds of objects, what kinds of expectations do they form about their movements? Do they expect all animals to move where they please? Do they expect all inanimate objects to stay put unless contacted by some other object?

In a series of studies carried out with my students, we presented wild rhesus monkeys and captive tamarins with a two-chambered box, separated by a partition with a hole at the bottom.[8] In every condition, an experimenter placed one object into one chamber, covered the box with a screen for a few seconds, removed the screen, and revealed the object inside the same chamber or the opposite one. When these monkeys saw an apple placed into one side, or a ball rolled in, they looked longer when the objects appeared in the opposite chamber than when they appeared in the same chamber. These are inanimate objects. They have no capacity to move on their own. Rhesus monkeys and tamarins are therefore surprised when a stationary apple or a human-propelled ball appear to move, on their own, to a different location. They showed the same pattern of looking when the experimenter placed a clay object with eyes in the center of the chamber and then, by means of magnets, invisibly caused it to move within the chamber. Thus, even though this object moved on its own from a stationary starting point—the definition of self-propelled—these monkeys were surprised to see it move to the adjacent chamber. However, when the experimenter placed a live animal—tree frog, mouse, hermit crab—into one chamber, both rhesus monkeys and tamarins looked as long when these animals appeared in the opposite chamber as when they appeared in the starting chamber. In the mind of a rhesus monkey or tamarin, therefore, living things hold a privileged position: Unlike nonliving things, animals can move where and when they want, or they can stay

put. Though self-propelled motion may provide a relevant cue, it isn't enough. When it comes to predicting an object's potential for trading places, these monkeys look for cues to animacy, hints that the thing they are looking at is alive, breathing, and capable of moving elsewhere.

Results from these experiments with monkeys lead to a potential difference with human infants. For our own species, a self-propelled object appears to provide sufficient cues to predicting an object's goals. The object in question can be as simple as a ball or a two-dimensional disk on a screen. A second experiment, however, suggests that monkeys may understand a corollary of principle 1: An inanimate object can only move if contacted by another object.

To further explore principle 1, the cognitive scientist Laurie Santos presented tamarins with one red and one blue train on a track. She then concealed the red train with a screen and launched the blue train. In one event, the blue train moved behind the screen, and, soon thereafter, the red train emerged from the other side of the screen. In the second event, the blue train only partially disappeared behind the screen, and soon thereafter the red train emerged. Thus, in the first event but not the second, the blue train made contact with the stationary red train. The second event is physically impossible, since the red train has no capacity to move on its own, and the blue train never made contact. Tamarins detected this impossibility, looking longer at the second than the first event. These results suggest that tamarins are equipped with the corollary to principle 1: Inanimate objects can't cause others to move without making contact.

Principle 2 builds on principle 1 by making goals an explicit part of the event:

PRINCIPLE 2: *If an object moves in a particular direction toward another object or location in space, the targeted direction picks out the object's goal.*

To check whether this principle is part of the mind's code, we can present an incongruous event, at least from the perspective of a normal human adult: an individual moves toward or attends to an object or location, and then heads off in a different direction or picks up a different object. Would, for example, an animal be surprised to see one individual

dash over to join another and then lie down and fall asleep? Would an animal be surprised to see another look toward a coconut but then reach for the banana? Woodward's experiments, described in chapter 4, addressed these exact questions. Babies watched as an experimenter looked at one of two objects on a stage and then reached either for this object or the other one. Infants looked longer when the experimenter reached for the unattended object. So, too, did cotton-top tamarins.[9]

It looks like we share principle 2 with at least one other animal. It is a principle of action with far-ranging moral implications, including our ability to detect rather perverse actions. When a mean-spirited parent teases her child by offering a toy that she can never reach, we perceive this as a moral infraction—as morally wrong. We recognize the perversity by recognizing the child's goal—grabbing the toy. Without a capacity to recognize goals and goal-directed behavior, we wouldn't have a category of morally perverse teasing. Animals, such as tamarins, have some of the requisite psychological machinery, even if they don't attribute moral perversity to teasing. Chimpanzees, however, apparently do make such attributions, and I will provide the evidence in a moment.

In 1984, while I was watching vervet monkeys in Kenya, I noticed a vervet infant who seemed to be irritated by something on its left thigh. It kept picking at this one spot. The infant's mother was some distance away, on the other side. All of a sudden, this infant leaped up in the air, bounding forward. Given the irritation, I assumed that something had poked or pinched the infant, causing her to leap up and forward. The infant's mother immediately dashed in to see what was wrong. But what did this vervet mother think? Did she assume that something pricked her infant, causing her to jump? Or was she puzzled at her child's apparent attempt to leap up and over an invisible barrier? Did she think that her child was acting irrationally? Principle 3 addresses this exact issue:

PRINCIPLE 3: *If an object moves flexibly, changing directions in response to environmentally relevant objects or events, then it is rational.*

Gergely and Csibra provided the key test of this principle with human infants, and the developmental psychologist Claudia Uller provided

a replication with infant chimpanzees.[10] Each chimpanzee sat in front of a television and watched as a small square moved toward and over a barrier, and then settled next to a large circle. Watching multiple reruns of this show, they then watched two new shows, each with the barrier removed. In one show, the square moved forward a bit, then arced up and down and then straight over to the circle; this mimicked the original trajectory, but from a human perspective—both adult and infant—it appears bizarre and irrational. In the second show, the square moved straight across to the circle—a perfectly rational action.

Chimpanzee infants looked longer at the irrational square, implying that they expected rational action from a geometric figure faced with a new environment. Principle 3 appears to be an evolutionarily ancient piece of the psychology of action—part of primate DNA.

When chimpanzees in certain parts of Africa engage in a grooming bout, one animal initiates the interaction by raising its arm. The partner, if interested in grooming, responds in kind, and then both lock hands in what is called hand-clasp grooming. Here, the timing of the initial arm-raising, followed by the mirrored action, sets up a contingent response. It sets up a social interaction. This is the core aspect of principle 4:

PRINCIPLE 4: *If one object's action is followed closely in time by a second object's action, the second object's action is perceived as a socially contingent response.*

Grooming is one form of cooperation, seen among a wide range of animals. Grooming bouts can be carried out sequentially, with large or small gaps between bouts, or at the same time. Other forms of cooperation involve similar sequential exchanges, including babysitting, alarm calling, and food sharing. To maintain these cooperative exchanges, animals must have some sense of contingency. They must have something like principle 4, even though no study to date has actually carried out an explicit experiment.

Where contingency arises for some animals is in games of cooperation, especially those that involve either two animals working together to achieve some common goal or some form of reciprocity. In a reciprocity game with tamarins, an experimenter trained one animal to play a unilateral

altruist strategy, always giving food to its opponent when the opportunity arose. The experimenter trained a second animal to play a unilateral defector strategy, never giving food to its opponent. If contingency matters, then when the untrained animals offer food to their trained opponents, food comes back from the unilateral altruist but never from the unilateral defector. All of the untrained animals cooperated with the altruist, which paid off handsomely, but not the defector. These observations, together with other experiments, provide evidence that contingency can play a central role in either stabilizing or breaking apart social relationships in animals.[11]

By their nature, all social animals have the skills to pick out the cooperators and cheaters, the kind and ferocious ones, the dominants and subordinates. Many animals form coalitions with trustworthy partners to gang up and defeat those higher up in the pecking order. Among monkeys and apes, when a dominant male moves near, meek subordinates spread their lips, baring their teeth in a display of submission that typically provides them with a protective shield—a passport against random acts of violence. From these observations, however, we don't gain the requisite insights into how these skills are acquired and how they are represented in the mind. We need to understand whether there is a key principle of action that determines how animals judge particular social interactions, assigning some to the category of helping and others to the category of harming. Are animals guided by principle 5?

PRINCIPLE 5: *If an object is self-propelled, goal-directed, and flexibly responsive to environmental constraints, then the object has the potential to cause harm or comfort to other like-minded objects.*

The most relevant experiments are ones by David Premack and the psychologist Josep Call, both focusing on an actor's goals and the relationship between actions and the personal nature of their consequences.[12]

Premack recruited his star chimpanzee, Sarah, for this complicated task. Following years of experiments, Sarah had trainers that she liked and ones that she disliked. Premack selected one of each for this experiment. For every test, Sarah first watched a videotape of a trainer attempting to

grab food that was just out of reach. Next, an experimenter handed her an envelope with three photographs. One showed a picture of the trainer using a proper action to solve the problem; for example, the trainer picked up a long stick to rake in the food. One showed a picture of the trainer using an improper action to solve the problem, such as picking up a short stick that didn't quite reach the food. And the final picture showed the trainer using a proper action but an irrelevant one to solve the problem; for example, the actor stood on a chair, a proper response to food hanging from the ceiling but an improper response to food placed out of reach on the ground, only accessible with a long stick. Would Sarah pick different actions depending upon whether she was watching the likable or unlikable trainer? If she was like us, she would want the likable trainer to succeed and the unlikable trainer to fail. If she was like us, she should pick the proper action for the likable trainer and the improper or irrelevant action for the unlikable trainer. If she was unlike us, she might just pick what happened next in the sequence, regardless of which trainer she was watching.

Sarah acted the way we would. In every condition, she picked the proper action for the likable trainer and either the improper or irrelevant action for the unlikable actor. These results suggest that chimpanzees recognize their own goal states, and can also represent the goals of others. And they can marry these representations of others with an assessment of their own emotions to choose actions that benefit some and potentially harm others. This capacity is central to morality, as it leads to the strategic use of cooperation with those whom we like and rejection of those whom we dislike.

In Call's studies, chimpanzees paired up with a human experimenter who controlled access to food—a highly desired grape. In some situations, the human experimenter cooperated, giving the chimpanzee a grape, and in other cases they didn't. At stake, however, was whether the chimpanzees would distinguish between actions that on the surface were similar but that differed in terms of the experimenter's underlying intentions or goals. Consider teasing versus clumsiness. In the teasing condition, Call held out a grape, moved it toward an opening in the partition, and then as soon as the chimpanzee reached for the grape, pulled it back. Clumsiness, in contrast, involved the same actions, except that Call accidentally dropped

the grape each time he moved it toward the opening. For both interactions, Call moved the grape toward the chimpanzee, and the chimpanzee never received the grape. If chimpanzees only cared about getting food—if they were merely consequentialists—then from their perspective, an experimenter who teased them would be no different from an experimenter who was clumsy. The consequence would be the same: no grape. If chimpanzees cared about why they did or didn't get food—if they cared about the means—then these interactions were different. Call would be morally perverse—going back to principle 2—in the teasing condition, but merely annoying in the clumsy case.

Chimpanzees see the difference between these two conditions. In response to teasing, and in contrast to clumsiness, they leave the testing arena earlier and show greater signs of frustration—banging on the window, aggressively calling. Whether they perceive the teaser as morally perverse is anyone's guess at present. So, too, is the question of whether they would generate the same attributions to an inanimate object performing the same actions, paralleling the studies of infants watching geometric shapes move on a television monitor.

Premack and Call's studies suggest that chimpanzees may have access to principle 5. Minimally, they appear to read beyond the surface features of action to the intentions and goals of the actor, using these as a foundation for distinguishing between those who help and those who harm. And, presumably, this is part of their psychological design, because selection favors capacities that ultimately feed self-interest, even if it is in the context of cooperating with others.

In contrast to the wealth of information on the human child's developing concepts about living and nonliving things, we know relatively little about the animal equivalent. This makes the evaluation of some of these principles less than satisfying when it comes to the moral domain, as we would ultimately like to understand how moral judgments shift as a function of the individual's understanding of life and death. As mentioned in chapter 4, the human child's understanding of death is a relatively late development. Do animals have anything like a concept of death? Is it like the young child's, anchored in facts about breathing and moving? Or is it richer, more theoretically informed, tapping notions of growth and reproduction? Unfortunately, we only have anecdotal observations to go on.

Some animals, such as the ants that Ed Wilson has described, clearly don't. When an ant dies, it is dragged out of the colony and deposited. But dead ants secrete oleic acid that, when placed on living ants, causes them to be deposited in the ant cemetery as well. For an ant, dead = oleic acid. For other species, the story is richer, but nonetheless unclear. Studies of elephants, monkeys, and apes suggest that individuals, especially mothers, go into a state of mourning upon losing their offspring. These observations indicate that the loss of a group mate causes a change in others' behavior, and, we presume, their emotional states. But it tells us little about their understanding of death, whether they have any expectations about this individual's future, whether they will ever return, or carry on in some altered state somewhere else. Yet without an understanding of their understanding of the life cycle, the connection between principles of action and moral significance remains tenuous.

WHO AM I?

While I was observing rhesus monkeys on Cayo Santiago, a BBC film crew paid me a visit. They were shooting a documentary on the emotions of animals and wanted to get some footage of rhesus monkey social life. They also wanted to set up a large mirror, to see what these monkeys would do. I warned them in advance that some of the more rambunctious juveniles and adult males may break it into small pieces. A large adult male soon kicked the mirror kung fu–style, shattering it into smithereens. End of film sequence.

As a group of rhesus looked on, we cleaned up whatever pieces we could find. Later that day, with the film crew gone, we saw four adult females walking around on three limbs, using the fourth to carry a small piece of mirror, periodically stopping to take a good, long look.[13] What were they seeing? What were they thinking? And why only the females? Given the size of the mirror, they couldn't possibly think that they were carrying someone else? And if not someone else, than whom other than "me"? Were the females self-absorbed beauty queens, trying to look their best for the macho boys?

A cottage industry of animal research has developed around the use of

mirrors to understand an animal's sense of self.[14] Charles Darwin initiated this approach with his studies of captive orangutans. But in 1970, more than one hundred years after Darwin, the comparative psychologist Gordon Gallup developed a more refined and informative method. Gallup provided chimpanzees with access to a full standing mirror and watched their behavior. Like the orangutans that Darwin had tested, the chimpanzees looked and made facial expressions at their mirror image, and also looked behind the mirror, as if they were trying to locate the individual inside, staring back. These behaviors did not lend themselves to a clean diagnosis. Gallup then took a further step. He anaesthetized each chimpanzee and, while they were unconscious, placed an odorless red-dye mark on one eyebrow and on one ear. Once they were conscious again, Gallup placed the mirror in front of them and watched. Immediately, the chimpanzees looked in the mirror and touched the dye-marked areas. This behavior can be interpreted in two ways. One, the chimpanzees figure out that when they move, the mirror image moves as well, in perfect synchrony. They conclude: "That's me." Two, they see the mirror image as another chimpanzee with red marks and wonder if they have the same. In both cases, the behavior reveals something to the staring individual about themselves. The second explanation seems unlikely, given the fact that once chimpanzees recognize the dye marks and touch them, they then proceed to use the mirror to look at previously unseen parts of their body. The mirror has become a tool.

A slightly different kind of experiment by the comparative psychologist Emil Menzel enriches our understanding of self-recognition in animals. Menzel wanted to understand whether chimpanzees and rhesus monkeys could use a video monitor of their arm to find a concealed target location. Rhesus monkeys never made it out of the initial training phase of the experiment, so there is nothing to report. Chimpanzees, in contrast, were not only able to use the video-monitor projection of their arm to find a concealed target, but were also able to reposition their arm when the image was inverted, and stopped reaching altogether when the monitor revealed a previously filmed version of their arm—in other words, when the real-time dynamics of their own arm moving stopped, they stopped moving as well. From the chimpanzee's behavior, we infer that it was thinking: "That's my arm on TV."

• Following Gallup's lead, several researchers wondered if their animals were also equipped with this ability, this sense of self, or whether they were as clueless as rhesus. One by one, as if Noah were administering some standardized test for admission onto the ark, experimenters marked parrots, pigeons, crows, elephants, dolphins, tamarins, macaques, baboons, orangutans, gorillas, and bonobos, showed them a mirror, and watched their response. And, one by one, most of these animals failed to touch the marked areas and failed to use the mirror to explore previously unseen private body parts. With the exception of dolphins, those that passed were close evolutionary relatives to the chimpanzees—orangutans and bonobos. Only one gorilla showed any evidence of mirror recognition— the language-trained and human-reared Koko, certainly not your average specimen.

Some researchers claim that chimpanzees, bonobos, and orangutans are special, while others claim that they are no more special than dolphins and gorillas, who also appear to pass this test. At the heart of this debate, however, are two uncontroversial points. First, not all animals will show evidence that they recognize their image in the mirror. Species differences could arise either because some animals lack this particular sense of self or because they are not particularly sensitive to changes in the visual domain, which would lead to detection of the dye marks. Instead, they may show greater capacities in other sensory modalities, such as hearing, smelling, or touching. For example, in a wide variety of species, especially songbirds, individuals respond differently to their own song played back from a speaker as opposed to the song of a familiar neighbor or an unfamiliar stranger; and in some songbirds, once an individual has acquired its own species-specific song, there are neurons that will only fire when the bird hears its own song. This suggests that, at the neural level, individual songbirds recognize their own song.

Second, the mirror test says nothing at all about what the individual thinks when it recognizes its reflection. We don't know if these individuals are appalled by their appearances, indifferent, or narcissistically mesmerized. We don't know what they know, how they feel about such knowledge—if they feel anything—and what they can do with it, assuming they can raise it to some level of awareness. One relevant piece of evidence comes from a set of experiments asking whether rhesus monkeys know when they are ignorant. To set up the problem, consider the movie *Memento*, a thriller that explores the nature of human memory. Although the actual story line is left intentionally vague, what is clear to everyone in the audience is that the protagonist can not remember any recent events. To aid recall, he tattoos key events onto his body, and posts sticky notes and photographs all over his room. He effectively offloads what would be stored in memory into an external videotape of his recent past. This trick works because the protagonist knows what he doesn't know. He is aware of his deficit and this allows him to counteract the problem.

The cognitive neuroscientist Robert Hampton[15] ran a series of experiments with rhesus monkeys designed to test whether they are aware of what they don't know. In one task, he presented subjects with a sample image, turned it off, and then offered a choice between a discrimination test or a pass. The test included four images, one of which was the same as in the sample. Hampton rewarded subjects for touching the matching image and punished them with a long lights-out period for picking any of the other, incorrect images. This is a standard matching-to-sample test, used in countless studies of nonhuman primates. Hampton's insightful twist on this standard was the pass option. On some proportion of trials, he gave subjects the option of passing up the opportunity to take the test trial and on the remaining proportion he forced them to respond. The idea was to give them the option of passing on the test when they were uncertain, perhaps because they had forgotten the details of the sample image. The key finding was beautifully simple: When Hampton forced rhesus to take the test, they did far worse than when they were in control of which test trials to take and which to pass over. Rhesus appear to recognize when they have forgotten, seeing ignorance as a deterrent to performance. This is one of the few clear pieces of evidence that animals know what they know, and can use this knowledge to aid action.

With studies like Hampton's, we can begin to see how to connect the different strands that constitute the animal's sense of self, especially the connection with their emotions and beliefs. Animals with these pieces in play would feel guilty about their own actions or expect guilt in another, recognize the difference between their own and another's beliefs, and use this knowledge to guide action and the judgment of another's actions. As the American philosopher Herbert Mead noted, organisms may only be

able to build a sense of self by recognizing the harmonious resonance between their own behavior and its mirrored reflection in another's behavior: "Any gesture by which the individual can himself be affected as others are affected, and which therefore tends to call out in him a response as it would call out in another, will serve as a mechanism for the construction of self."[16]

CROCODILE TEARS

In 2002, the Takara Corporation in Japan released Bowlingual, a digital device that translates dog barks, growls, and squeals into Japanese or English. The press release described the device as an "animal emotion analysis system," designed to "fulfill the realization of real communications between humans and animals." The device is rather simple, taking only three steps to deliver a translation. Step one: record the dog's vocalizations. Step two: analyze its acoustic morphology. Step three: convert the acoustic signal into one of six categories corresponding to different emotional states. If the analysis detects frustration in the dog's voice, perhaps because Rover wants to go outside while owner Bob is couched up watching the Superbowl, Bowlingual spits out phrases such as "I've had enough of this!" or "You're ticking me off!" If the analysis detects sadness, Bowlingual throws back "I'm bored" or "I'm sad."

The device was an immediate success. *Time* magazine dubbed it one of the best inventions of 2002, and the spoofy scientific magazine *Annals of Improbable Research* awarded Bowlingual its Ig Nobel Peace Prize for inspiring harmony between species. I imagine that for some pet owners, Bowlingual's decoding takes all the pleasure out of living with another species. As the American political commentator Andy Rooney once said, "If dogs could talk, it would take all the fun out of owning one."

But for those who have bought into Bowlingual, there is some financial investment in the company's promise of emotional decoding. But is this what you get for your money? Is there an acoustic signature of emotional frustration that maps onto behavioral frustration? Can sadness be plucked from the waveform? Many biologists, myself included, have spent significant parts of our careers trying to decode what animals are

saying, and none of us feels as comfortable as the Takara Corporation in labeling each acoustic signal with a descriptive label that is as clear as the ones Bowlingual generates. Perhaps scientists are too cautious, or perhaps they have struggled to find coherent explanations of what animals feel when they communicate. The staff at Takara Corporation has a different mission, presumably driven by money as opposed to accuracy. Within the first few months of launching their product, they had sold 30,000 within Japan, with a price tag of $220 per unit. Sales skyrocketed to 300,000 by March of 2003, with comparable sales on the international market.

But what do we learn about animal emotions from the work behind Bowlingual? Is it a cute gimmick or something more? Ever since Darwin, it has been clear that animals have emotions. Who could doubt that a growling dog is angry, a purring cat content, or a screaming monkey afraid? Controversy arises, however, in assessing whether the words we use to describe these emotions actually reflect the animal's experience, whether there are emotions that other animals experience but we don't— and vice versa—and whether animals make decisions that are not only fueled by the emotions but reliant upon them.[17] Here, I use this controversy to think about how our current understanding of animal emotions contributes to our understanding of their cooperative and competitive acts, behaviors that are guided by principles essential to the healthy functioning of any social system.

Consider fear, an emotional state that is apparently experienced by many animals, presumably because of its adaptive role in avoiding predators and competitors.[18] The logic of emotions, like the logic of our conceptual knowledge and systems for learning, may also be domain-specific. Fear of snakes is different from fear of heights or impending pain. As the social psychologist Susan Mineka has demonstrated, humans and other animals are equipped with a kind of mental readiness to respond with fear to snakes. If a group of rhesus monkeys with no snake experience watches an experienced group express fear toward the snake, the observers will readily absorb this fear, responding with alarm the next time they confront the snake. In contrast, if a naïve group of rhesus watches other rhesus show fear toward a bed of flowers, the fear doesn't spread; the next time they confront a bed of flowers, there is no fear at all. It would take a lot more to convince the primate mind that flowers count. And even if

they could be convinced, the process of associating flowers with fear would be different from the spontaneous fear that emerges in the context of seeing a snake. This kind of fear, also exhibited in humans with or without extensive experience with snakes, is different from anxiety.[19]

In contrast with monkeys, apes, and humans, rats do not have a characteristic facial expression for fear. They do, however, have both a freezing and a withdrawal response to things that they find threatening. Rats, monkeys, apes, and humans all show a cascade of hormonal and neural changes when frightened. For example, when frightened by an aversive event—a loud sound or visual cue previously associated with a physically painful experience—all mammals show activation of the amygdala.

Due to the overlap in physiological and behavioral responses, many argue that rats, monkeys, apes and humans experience fear. Others disagree, arguing instead that the actual experience is different, even if there are parallels in behavior and some aspects of the physiology. For example, the developmental psychologist Jerome Kagan argues that "One good reason for distinguishing between the state following a painful shock in rats and in humans is that the latter have a much larger frontal lobe. When humans hear a tone that had been associated with electric shock, the frontal lobes are activated and the person quickly acquires control of the biological signs of fear after only two exposures to the tone. That phenomenon could not occur in rats."[20] Although Kagan may be right, his comment concerning species differences depends on two untested assumptions: The size of the frontal lobe is crucial for the experience of fear, and the speed with which we acquire an association between tone and fear matters. The fact that fear activates the frontal lobes in humans, and not in rats, is interesting in terms of what areas of the brain are involved. But from a description of brain areas, it is impermissible to jump over and assume that the processing and experience are different. Rats may process the situation in a different part of the brain, but then experience the emotion in the same way we do. The issue of speed runs into a different problem. Although we may form the association between tone and fear faster than a rat or monkey, once acquired, each species may experience fear in precisely the same way. What is different is the learning mechanism that facilitates making the association, and this may, in fact, be due to our

larger frontal lobes. But this interpretation shifts the argument from species differences in emotion to species differences in learning.

Kagan is absolutely correct in pointing out that from the rat's behavior we must not leap into an inference concerning its subjective experiences—its feelings. When I say that I have a fear of heights, you certainly can't understand my experience if you don't have a fear of heights, and even if you do, you can't understand exactly what it is like for *me* to feel such fear. However, because members of our species share a common neural and physiological substrate, some aspects of our experience will be shared. Consequently, when you say that you are afraid of heights, I have a general understanding of what you mean. I also know, because I am a native speaker of your language, that when you say "fear," it refers to a distinctive kind of emotion. In the case of animals, we simply don't have access to all this information, nor is it necessarily reasonable to make the same kinds of assumptions. The same concerns hold for human infants. When an animal or human infant freezes, presents an increase in heart rate and the stress hormone cortisol, and then heads in the opposite direction from the apparent triggering event, we reasonably call these the signatures of fear; these are, after all, the same kinds of responses that human adults often make when they are afraid. These signatures indicate that some part of the brain has made an evaluative judgment about the situation that causes fleeing or fighting. But what we don't know is what fear feels like to each individual when they are in the throes of the experience. Let's put this hard question to the side, and consider instead how the perception of an event—imagined, anticipated, or real—triggers an emotion, and, on occasion, a follow-up action.

In socially living animals, either in the wild or in captivity, emotions undercut much of their daily life. Animals engage in political strategizing, attempting to climb the social hierarchy or avoid dropping any further within it. Climbing requires motivation, risk-taking, and aggression, while maintaining the current status quo requires sending signals of submission and fear to others higher up in the hierarchy. Mothers, and sometimes fathers, must contend with weaning their offspring—an often trying experience, as the infants' capacity to pester, torment, and manipulate are unmatched. To cooperate, individuals have to muster motivation and

reconciliation

trust. Fights will happen, perhaps mediated by feelings of revenge and a thirst for retaliation. But staying angry at someone that you have to live with isn't productive. Making peace is better.

Some of the most revealing work in this area comes from the detailed observations by the biologist Frans de Waal. Beginning with his classic book *Chimpanzee Politics*, de Waal has helped show the complexities of primate social life, highlighting the role that emotions may play in fueling competition and stabilizing cooperation in the service of preserving peace.

Following aggressive conflict, many nonhuman primates—and some nonprimate species, including dolphins, goats, and hyenas—attempt to reconcile their differences by engaging in a variety of peace offerings, ranging from hugs, kisses, and testicle-touching to grooming and the exchange of food.[21] Conflict is associated with stress, reconciliation with the reduction of stress. Researchers measure stress in animals by watching their behavior and recording physiological markers, including heart rate and levels of blood cortisol. Though stress serves an adaptive function, placing individuals in a ready state for action, prolonged stress compromises the immune system and can lead to selective neural death and, ultimately, early mortality. Among rhesus monkeys and baboons, heart rate and cortisol levels skyrocket following aggressive conflict, and remain above normal resting levels for several minutes. But when conflict is followed by a peace offering, heart rate and cortisol levels drop, as do accompanying behavioral correlates of stress. Though we don't know whether the experience of stress in monkeys, apes, and humans is the same, there are many behavioral and physiological parallels, including convergent changes following reconciliation.

The broad distribution of reconciliation among mammals is accompanied by important differences between species in how, when, and how often they do it. This tells an interesting story about the biology of reconciliation, especially its development and plasticity. Some species, such as the despotic rhesus monkey, rarely use reconciliation as a response to postconflict stress and ambiguity. Rhesus are much more likely to redirect aggression: If rhesus A beats up rhesus B, B is more likely to go and pound rhesus C than to hug rhesus A. In contrast, the egalitarian and closely related stump-tailed macaque is more likely to hug than fight. To determine

whether these differences are part of each species' innate repertoire, and unlikely to change even in a different environment, de Waal and his colleagues carried out an experiment involving some baby swapping. Rather than have rhesus grow up in their native environment, these youngsters were transported at an early age to a stump-tailed macaque colony. Would these young rhesus carry the flag of their despotic heritage or bend at the will of an egalitarian society? They bent. Rhesus monkeys reconciled their differences using stump-tail gestures. When they returned home to their native environment, these rhesus monkeys preserved their peacenik style, using reconciliatory gestures to manage conflict. Bottom line: genes enable certain species to reconcile their differences, but details of the local society guide whether they reconcile, how often, in what contexts, and with what techniques.

The work on reconciliation shows that emotions play a central role in the maintenance and guidance of certain social norms, even if the more immediate goal is to reduce stress and violence. If we had a simple method to evaluate primate judgments, we might say that the Humean creature fuels its judgment concerning what constitutes a permissible or possibly even obligatory situation for reconciliation. If a chimpanzee watched a film of two individuals fighting, and then saw a follow-up sequence in which they did or did not reconcile, what would be the more surprising case? What would they expect? What counts as a violation or social transgression? Although emotions play some role here, we are also left with the same dilemma that confronted our account of human judgment. To evaluate the interaction, chimpanzees must also recognize it as a case of aggression, assess whether harm was intended as a direct or indirect consequence, evaluate the time elapsed postconflict, and consider the local society's expectation with respect to the form of reconciliation. Given this calculation, carried out without emotional input, a chimpanzee might judge whether reconciliation is permissible or obligatory. The Rawlsian creature is back. Unfortunately, few researchers in this field have looked at reconciliation with respect to this kind of appraisal mechanism,[22] leaving the door open to at least two different accounts: Both emotional and action analyses drive their expectations, or emotions follow from the analysis of action. Whichever way this turns out, there is one

obvious conclusion: In species with reconciliation, as well as other dyadic or even triadic social relationships, there are principles of action in play that generate expectations about how animals ought to behave.

At present, there are two competing explanations for why reconciliation evolved as a form of conflict resolution in animals. One possibility is that selection favored reconciliation because of its role in preserving long-term, valuable social relationships. A second is that selection favored reconciliation because it enables individuals to send benign signals of intent, designed to reestablish cooperative alliances for short-term resource gains. Both explanations put a premium on the value of the relationship, either for its own sake or for the immediate resources it affords. Here, then, is a way of marrying the Humean creature with values, and some measure of utility. We can ask how much are such relationships worth? Do animals feel that social relationships are part of their natural-born rights? How hard are they willing to work for them? Is depriving an animal of a social relationship a violation of an implicit moral code?

A way to get at these vexing questions about what really matters to animals comes from an unexpected source: a series of experiments explicitly designed to address questions of animal welfare and rights. In the 1980s, the ethologist Marianne Dawkins and her students developed a brilliant line of experiments based on a simple economics model.[23] The work starts from the premise that for the near future, our species will keep other species in captivity so that we can eat them or use them for some biomedical purpose. Some readers will vehemently disagree with this policy, but the fact remains that many humans enjoy eating animals and, for a variety of human ailments, research on animals provides the only current hope for a remedy. Given that we are going to keep animals in captivity, the only humane thing to do is to treat them with respect, and give them what they need. We can figure out what they need by studying what they do and what they have in their native environments. Finally, we can use what we learn from these observations to create an economy in which animals can work—pay—for what they want, and thus, presumably, for what they need; we call this a closed economy, because there are only a set number of products that an individual can purchase.

In one of the first studies to adopt this approach, Dawkins explored what domestic chickens need. The experiment was motivated by a decision

from the British government stating that, due to rising costs, chickens could no longer be supplied with wood shavings on the bottom of their cages. Dawkins argued that chickens need such shavings because it allows them to carry out their species-typical scratching behavior. Dawkins placed a hen on one side of a two-chambered box, separated by a door. The only difference between the two chambers was that one had wood shavings on the floor and the other was bare. Hens placed on the side with shavings stayed put, while hens placed on the bare side immediately moved over to the side with shavings. Next, Dawkins made it more difficult for hens to move from one side to the other by increasing the tension on the door's spring. Although the costs of moving increased dramatically, hens placed on the bare side rammed into the door, eventually making their way to the wood shavings. Chickens not only want wood shavings, they need them.

A similarly designed study examined what mink want, in order to evaluate their housing conditions in fur farms. Each mink started in a standard cage, but with an opportunity to upgrade by choosing different commodities, each placed behind a different door. Behind door 1, mink found a larger cage; behind door 2, a second nest site; door 3, a raised platform; door 4, a tunnel; door 5, some toys; and, behind door 6, a water-filled pool.

Mink consistently opened door 6, content with the opportunity to bask in the water. And, like Dawkins's chickens, mink paid the high price of admission to water by ramming through the heavy, spring-loaded door. Most significant, from the perspective of welfare and our understanding of the mink's emotions and values, mink denied access to water pools were physiologically stressed, almost to the level of mink denied access to food. If their evolved right to live with and in water is taken away, these animals are continuously stressed. Continuously stressed animals develop compromised immune systems. Animals with compromised immune systems are more susceptible to disease, and therefore more likely to die prematurely. That seems unfair and wrong.

What do mink want? Water pools. Why? Because in nature, mink spend a considerable amount of time in the water. Water is a necessary commodity.

From Dawkins's initial insight, designed to infuse objectivity into the

often-subjective debates about animal welfare, we gain a new understanding of how animal emotions connect with animal values. We learn what animals need, what they will fight for, and how selection molds a relationship between the value of a commodity and their motivation to work for it.

Crocodiles don't shed tears, and elephants don't weep. No animal expresses its sorrow by turning on the eye faucets. This is a uniquely human expression. But underlying this human specialization are a heart and mind that share many commonalities with other animals. And in this sense, the Humean creature has an ancient evolutionary heritage. That it has such a legacy does not imply a static psychological system that is no different today than it was when we diverged some 6–7 million years ago from a chimpanzee-like ancestor. How we experience emotions must, in some way, differ from how animals experience emotions. But so, too, must chimpanzee emotions differ from elephant emotions, which must differ from crocodile emotions, which must differ from ant emotions. The main point here is that whatever emotions animals have, they are involved in individual action and the evaluation of others' actions.

NATURAL TELEPATHY

In the 1960s, the computer scientist John Conway developed a program called Life. Though built from a few simple rules, it provided an elegant example of how chaos can morph into order. The game is played on a grid. Each cell has eight neighboring cells, and each cell is either alive or dead. Only three rules bring this static grid to life:

1. If a cell has one or no living neighbors, it dies of loneliness
2. If a cell has four or more neighbors, it dies of overcrowding
3. Whenever an empty square has exactly three living neighbors, a new cell is born

From a few randomly filled-in grids, we move quickly into a series of organized clusters of life, as some cells die and others are born.

The standard game of Life involves extremely simple creatures, perhaps mindless, guided by three rules. These creatures have no social relationships.

What happens if we input social relationships into the game of life? Imagine a game involving a fictional species with two distinctive types. Let's call them B and M, for Behaviorist and Mentalist, respectively. These two types look the same on the outside, but are different on the inside. Bs make decisions about social interactions and relationships using only their prior experiences. By accumulating data, they spend more time with some than others. They use simple statistics to classify the population into friend or foe. Their prior associations define what they do to and with other group members. Ms make use of experience to guide their interactions, but go one step further. They make inferences about what is unobservable: the beliefs, desires, and intentions of other group members. They are mind readers, using information about what other individuals can or cannot see, or what they do or do not know, to predict what they will do next. Ms make predictions about behavior in the absence of having experienced behavioral interactions with others. Where someone is looking represents a proxy for what that individual knows. What they can't see, they can't know, assuming that the senses of hearing, touching, or smelling are out of commission. Ms can use their knowledge of what others know to teach and to deceive. This ability to infer what can't be seen means the Ms are better behavior readers because they go deeper into what behavior implies about believing and knowing.

Now imagine a simulation on the grid of Life. Here, we are looking for not only a shift from chaos to order, but an insight into who will win out and why. If both Bs and Ms reproduce, who will make more babies, winning the Darwinian footrace that is measured in terms of genetic prosperity? If one wins the reproductive competition, then there is room for selection to operate, favoring one and weeding out the other. Selection will favor the best design given the environmental circumstances. Ms are faster and more insightful than the Bs, and they are up to the challenge of both a novel habitat and completely novel social interactions. Bs sit around and wait for more data. They rely on highly familiar cues for deciding the next move. As a result, Bs make silly mistakes, failing to distinguish two actions that look the same but differ, because one was done intentionally and the other accidentally. The chaotic population of Bs and Ms will end up as an orderly grid of Ms. Individuals that can predict what is going to happen before it happens are like good chess players: They are

several steps ahead and thus can manipulate their opponents by seeing where others will fail or succeed. In the Darwinian competition of life, Ms live, Bs die.

Up until a few years ago, most essays on human evolution concluded that we are the only Ms; all other animals are Bs. While we are uniquely mind readers, everyone else is a mere behavior reader. In my book *Wild Minds* (2000), I echoed the consensus view that animals fail to make inferences about others' ". . . beliefs, desires and intentions—they lack a theory of mind." I followed this comment up, however, with a more cautionary note, based partially on wishful thinking and partially on an insider's knowledge of new experiments by a young graduate student: "We must be cautious about this conclusion, however, given the relatively thin set of findings, weak methods, and incomplete sampling of species and individuals within a species." Here I want to capture the current state of play in a rapidly changing field, including what we know and how it bears on the central ideas that David Premack set in play about twenty-five years ago.[24] Does any animal, other than the human animal, move beyond behavior and into the minds of other individuals? If so, what kinds of psychological states can animals read, using this information to predict behavior before it happens?

Two sets of experiments, one on macaques and the other on chimpanzees, dominated the comparative landscape up until the end of the millennium.[25] Both led to the same conclusion: Animals, even chimpanzees, are strict Behaviorists! Dorothy Cheney and Robert Seyfarth showed that macaque mothers expressed the same level of alarm when their offspring could see an oncoming predator as when they could not. In the context of predation, ignorance is not bliss. But macaque mothers acted as if there was no difference. They failed to distinguish between an ignorant and knowledgeable infant. They also failed to take into account what infants could see and, therefore, what they would know. And the same story plays out in studies of other monkey species: In baboons living on the savannas of Botswana, for example, mothers don't call back to their distressed offspring, even though this would provide the babies with explicit information that the mothers are aware of their plight.

The anthropologist Daniel Povinelli presented comparable findings based on a series of studies of chimpanzees. In the general setup, a chim-

panzee entered a test room and, for each condition, had an opportunity to beg for food from one of two experimenters. In each condition, one experimenter could see the begging chimpanzee and the other could not. For example, one experimenter faced the chimpanzee while the other turned his back; one experimenter looked off to the side while the other looked straight ahead; one had a blindfold on his eyes while the other had a blindfold on his mouth; and one had a bucket on his head while the other held the bucket to the side of his head. With the possible exception of one person turned around while the other faced forward, the chimpanzee's begging behavior was random, even with massive amounts of training.[26] They were as likely to beg from someone who could see as from someone who could not. Like the macaques, these chimpanzees were as likely to make a request from an ignorant experimenter as a knowledgeable one. Macaques and chimpanzees are mind-blind.

There are at least two reasons why these findings seemed paradoxical at the time.[27] First, there was a mountain of anecdotal evidence from wild and captive monkeys and apes showing that they are sensitive to where someone is looking; their sensitivity shows up in what biologists describe as tactical deception, the strategic manipulation of another's access to information for some self-serving benefit. For example, low-ranking animals sneak copulations or pinch a piece of food when the dominant alpha isn't looking. Though everyone acknowledges the need for caution when interpreting these single-observation cases, piling them up amounts to an impressive set of observations, raising the possibility that nonhuman primates deceive by taking into account what others can see and potentially know. Other work on plovers and jays suggested that these birds consider where someone is looking when they engage in concealment behavior, hiding their nest in the case of plovers and hiding a food stash in the case of jays. Second, several studies showed that monkeys and apes attend to where others are looking, and can use this information to pick out what someone is looking at. For example, if a chimpanzee enters a test room and sees a human experimenter staring up at the ceiling, he will immediately look up to the same area; seeing nothing at all on the ceiling, he will then glance back at the experimenter to recheck the direction of gaze and then look up again. Given the combination of anecdotal evidence on deception and experimental work on reading visual perspective, there was a

growing tension in the field between the believers and the nonbelievers. This tension helped set the stage for Brian Hare—the unnamed graduate student from a few pages back.[28]

Hare's insight was simple. Chimpanzees in the wild compete more often than they cooperate. Their competitive skills have evolved to handle other chimpanzees who have similar interests in limited resources, including food and potential mates. Povinelli's experiments, in contrast, involved cooperation, and, in particular, cooperation across the species' divide: between a chimpanzee and a human. Might chimpanzees recognize the relationship between seeing and knowing if they had to compete with each other for access to food?

Hare's experiments involved a competitive task between two chimpanzees of different rank. Each separate experimental setup or condition explored the same question, but from different angles: Would these two competitors use information about seeing to make inferences about knowing, and then use this information to guide the next competitive move? Each condition imposed different constraints on what either the subordinate, dominant, or both could see. For each condition, the subordinate sat in one room, the dominant in an adjacent one, and a test room between them. When the dominant and subordinate had the same visual access to the available food in the test room, the subordinate stayed put and the dominant ran out and grabbed it all. But when the subordinate could see hidden food that the dominant could not, the subordinate headed straight for it. For example, in one condition, Hare set out two opaque barriers in the center test room. While the subordinate watched, and the dominant looked away, he concealed one banana on the subordinate's side of the barrier. Although subordinate chimpanzees typically avoid conflict over food when dominants are nearby, in this condition, they beelined to the hidden piece of food, taking advantage of their exclusive visual access. These results, together with several other conditions, show that chimpanzees can use seeing to outcompete others. They imply that chimpanzees can use seeing as a proxy for knowing.

These results are interesting on another level. The patterns observed do not reflect individual personalities, but rather, the relative ranks of each individual in the pairing. In some contests, an individual held the dominant position, and in other contests the subordinate position. Their

behavior changed as a function of their current ranking. For example, in a condition in which Hare placed one banana in the open and one hidden behind an opaque screen, individuals changed their strategies depending upon their relative rank: When subordinate, they first moved to the hidden banana and then to the visible piece, whereas when they played dominant, they moved to the visible banana first and then to the hidden piece. What determines how an individual competes for food is not how his opponent behaves, but what his opponent can see and therefore know about the current arena of competition.

Hare's results opened the floodgates to further studies of chimpanzees and other species—including monkeys, apes, jays, and ravens—pushing the logic of the initial experiments, especially the use of natural, untrained behavior.[29] For example, studies of captive chimpanzees and wild rhesus monkeys explored the relationship between seeing and knowing by using Povinelli's original design but with one critical change: Instead of subjects cooperating with a human experimenter, they competed. Consider the rhesus results carried out by the psychologists Jonathan Flombaum and Laurie Santos as they more directly parallel Povinelli's design, and also push the evidence further back in evolutionary time to a species that diverged some 30 million years ago from the branch that ultimately became human. Two experimenters approached a lone rhesus monkey on the island of Cayo Santiago, separated by a few feet, and then each placed a white platform with a grape on top next to his feet. For each condition, one experimenter kept an eye on the subject, while the other either looked away or couldn't see due to an opaque barrier. For all conditions, rhesus monkeys selectively snuck food away from the experimenter who couldn't see them.

Animals as distantly related as birds and primates use seeing as a proxy for knowing. These animals have evolved the ability to go deep, reading minds to predict behavior.

These new results on mind reading are only the beginning. There are controversies here and there, as well as further pieces of the story to map out. We need to understand in what ways mind reading in humans and other animals are similar and different; both similarities and differences bear on the extent to which animals can recruit an understanding of beliefs, desires, and intentions to make judgments of moral importance.

N. Selection favors bounded R,
+ context-specific R.

One way in which humans and other animals may differ is in the extent to which they deploy their mind-reading capacities across different contexts. Across several studies of chimpanzees, results show that individuals successfully use information about what another individual knows and intends to guide competitive interactions, while failing to use the same information to guide cooperative interactions.[30] From a human perspective, these results are puzzling. If I know that you are ignorant about the location of a hidden piece of food, I can both outcompete you because of your ignorance or guide you to the right location in order to facilitate cooperation. The context is irrelevant, because our capacity to mind-read is more general and abstract. How shall we interpret the results on chimpanzees?

Several students of animal behavior have noted that selection appears to have favored highly context-specific adaptations, designed to solve a small range of problems. This has led to the idea that animals have laser-beam intelligence while we have a floodlight of brilliance. One explanation of the chimpanzee results is that their capacity for mind reading is different from ours, able to use seeing as a proxy for knowing when in the heat of competition, but not in other contexts. This specialization is akin to the honeybee's famous dance-language. When first described by the ethologist and Nobel laureate Karl von Frisch, it was described as a language, because the dance was symbolic, providing detailed information about the distance, direction, and location of food displaced in time and space—all characteristics of words, and the more general capacity to refer to objects and events in the world. As it turned out, however, the honeybees' capacity lacked generality. It was remarkably specific, restricted to food and nothing else besides food. Although it is conceivable that honeybees have little else to talk about, further work by von Frisch and other students of bee biology have noted the rich complexity of their social lives. Bees have much to talk about, but don't, at least not with the referential precision of their foraging dance. Their communication system is an example of laser-beam intelligence. The social psychology of the chimpanzee may be another example.

There is an alternative explanation for the chimpanzee results, one that takes us back to chapter 4, cheater detection, and the Wason card-selection task.[31] To recap, here is the argument that Cosmides and Tooby

used to both motivate and interpret their results. Humans have been selected to solve problems involving social contracts as these are the kinds of problems that we evolved to solve in our hunter-gatherer past. In contrast, we did not evolve to solve abstract, socially detached problems of logic. Proof comes from human performance on the standard Wason logic test and Cosmides and Tooby social-contract version. We draw the correct inferences when the logic is translated into the language of a social contract, but not when it is in a more pure, unadulterated form. There is a context effect that plays on our ability to draw logical inferences. From these results, we do not conclude that humans are like honeybees, with a laser-beam intelligence for generating logical inferences. Rather, we conclude that context can sometimes uncover masked abilities. A similar explanation is possible for the chimpanzee results. It is only in the context of competitive interactions that we can unmask what lies behind the chimpanzees' eyes. Given the chimpanzees' prowess for cooperation in the wild, my guess is that it is only a matter of time before someone reveals comparable cases of mind reading in this context as well.

If we put together all of the results on mind reading in animals, the conclusion seems clear: We are not uniquely in possession of this capacity. Premack's early intuitions about the chimpanzee's theory of mind were right. How far does this capacity reach in animals? Do animals recognize that others can have false beliefs? Do animals recognize the difference between accidents, mistakes, and informed choices? At this point, it is too early to say. In the absence of such information, however, we can't say how rich or impoverished the animal mind is with respect to judgments of others' social actions. We can't say whether violations of social norms are judged on the basis of consequences or the causes that drive them. There is an urgent need to know more about what animals know about each other.

WEIGHTING WAITING

Many birds and rodents stash food in secure places for weeks or even months, and then use their razor-sharp memories to return to these hiding places for a feast. Many spiders, fish, and cats sit for long periods of time,

quietly watching a parade of prey before pouncing on an inattentive individual. A wide variety of primate species spend considerable time peeling, stripping, or cracking into highly protected fruits before reaping the rewards of their efforts with a delicious slice of flesh. And most animals face the general problem of whether to stay with the current patch of food or move on to greener or fruitier or meatier pastures. For each of these cases, individuals must fend off the temptation to feed an immediate desire, waiting for a more profitable but delayed return. They must delay gratification. It looks like evolution equipped animals with a healthy dose of self-control.

Foraging problems, such as those mentioned above, involve decision-making. Assume, as is standard in the field of animal behavior, that natural selection has designed animals to maximize foraging returns, converting energy into babies. In absolute terms, a small amount of food is worth less than a large amount of food; ditto for low- and high-quality food items. Where things get interesting is when the small or low-quality food item is available immediately whereas the large or high-quality food item is available at some point in the future. For example, imagine that a leopard sees a small, lame, juvenile gazelle only a few feet away, but a large, fat, and healthy adult female one hundred feet away. If the leopard attacks the juvenile, it will succeed and feed immediately. If it passes up this opportunity and hunts the bigger adult, it will take more time and energy but the returns will be greater. The central problem is how time influences this choice process. Waiting for a larger or more valuable food item is risky: The probability that food will be available in the future decreases over time, as other competitors may jump in and snatch it away, or the vagaries of climatic events may damage it. We want to understand the kinds of calculations animals make as they look at the trade-offs of taking something immediately as opposed to waiting. How far does the value of a food packet sink with time? Are there some trade-offs that no animal would ever contemplate, thinking that no matter how tasty a piece of food might be now, it would never have the same value as the life of an offspring? If animals show limited self-control, acting impulsively in the face of temptation, then they will break down when called upon to follow social norms. They will succumb to self-interest in the face of helping another. The short-term gains to self outweigh the potentially long-term but delayed gains from cooperation and being nice to others.

The relationship between value and time falls under the general topic of temporal discounting: the longer the delay to accessing the resource, the lower its value. There is a vast literature on discounting in rats and pigeons, and a smaller set of studies in less traditional laboratory animals, such as starlings, jays, tamarins, marmosets, and macaques.[32] Paralleling studies in humans, the central question is: How does the value of an item or action change as a function of time? Economists tend to think of the relationship between value and time as an exponential curve: The subjective value of a reward some time in the future decreases at a constant rate. This decay is therefore a measure of risk, of potentially losing everything by waiting for the larger reward. In contrast, students of human psychology and animal behavior tend to think of this relationship as a hyperbolic curve. Like the exponential model, there is a trade-off between subjective value and time, but with two distinctive differences: value is inversely proportional to time delay, and preference reversals arise when the time delay to both rewards stretches out into the future. Preference reversals are real in humans, a fact that annoys economists with a bent toward rational choice, but delights psychologists interested in the basis of subjective preferences. The exponential model can't explain why a human who prefers $10 today over $11 tomorrow would flip this preference when offered $10 in thirty days and $11 in thirty-one days. Since the difference in delay is the same and the monetary rewards are the same, the preference should be the same. The hyperbolic model, in contrast, predicts context effects such that rewards dispensed in the future have an inherently different subjective feel than rewards delivered in the immediate present. Humans flip-flop their preferences depending on time. The hyperbolic model predicts this pattern.

Give pigeons the choice between one and ten food pellets. They consistently pick ten; so will every other animal. Now, make pigeons work for their food. If they peck the left button, they immediately get one pellet, whereas if they peck the right button, they get ten pellets later. If "later" is much more than a few seconds, pigeons will consistently peck the left button for one pellet. They can't resist. The value of one piece of food drops precipitously after a short wait. Their impulsivity persists as long as there is a good-sized difference between the small and the large and there is some waiting period for the large and little or none for the small. In species as different as pigeons, rats, tamarins, and macaques, the ability to

wait for a larger reward is on the order of seconds. Humans given a similar task will wait for hours and even days. No contest. When it comes to patience, we are the paragon of animals.[33]

In some sense, preference for the smaller immediate quantity is irrational. If selection favors long-term gains, because these impact most on survival and reproduction, then animals should wait it out. As the behavioral ecologist Alex Kacelnik rightly points out, when a pattern of behavior is observed in a wide variety of species, and when the consequences of this behavior appear to go against the ultimate goal of maximizing genetic fitness, it is high time for evolutionarily minded scientists to figure out why. The experimental economist Ernst Fehr offers this explanation:

> . . . throughout evolutionary history, future rewards have been uncertain. An animal foraging for food may be interrupted, or, in the case of reproductive opportunities, die before it is successful. For humans, the promise of future rewards may be broken. And if the risk faced by a person varies over time, he or she applies various discounts to future events and so behaves inconsistently.[34]

This account implies that animals, humans included, are nonoptimal, failing to maximize their potential intake because they are chained to the ghost of uncertainty. But sometimes what appears to be a maladaptive solution may represent an appropriate solution under different circumstances. Consider, for example, the typical laboratory task offered to pigeons and rats, and what the naïve animal must learn. At first, the individual wanders aimlessly around his cage, doing nothing much at all. Eventually, it stumbles onto a lever, pecks or presses it, and something happens either immediately or with some delay. Since the causal force of an action is greatest with short delays, contacting the lever associated with no or little delay is immediately most effective. Consequently, there will be a bias in the learning phase to make contact with the lever associated with the small immediate reward. And this bias maps onto the natural foraging behavior of most animals in most feeding contexts. In nature, foraging decisions almost never entail an action followed by passive waiting; in those cases where it occurs, such as the food storing of birds and rodents, there is an entire period devoted to storing and then a long follow-up period devoted to waiting prior

to retrieval. We can therefore explain the bias to grab the more immediate small reward by the fact that there is a more transparent relationship between grabbing and getting a reward. Learning to wait for some abstract period in the absence of doing anything is unnatural—at odds with the biologically engineered machinery for learning.

To circumvent some of these problems, some students of animal behavior have followed Kacelnik's lead, using nonstandard laboratory animals to explore decision-making under more realistic conditions. In particular, whereas the traditional laboratory experiments on discounting give animals choices between two options and explore how waiting impacts upon choice, the more realistic tasks translate waiting into a behavioral measure of expenditure, using insights from subjects' native environments to establish appropriate experimental conditions. When animals forage, rarely do they just sit still and wait for food to arrive on a silver platter; the exceptions are the sit-and-wait predators. Most species walk, run, fly, scratch, peel, and pry in order to eat. Foraging animals *behave*. In one of Kacelnik's studies, he gave starlings a choice between walking or flying for a reward; each activity was associated with a particular energetic cost and return rate for food, with flying associated with higher costs but better returns. Starlings followed a hyperbolic pattern that maximized intake per unit of time. In Kacelnik's terms, starlings deployed a rational strategy, given the constraints.

Taking into account a species ecology can also reveal how selection molds different patience functions. The behavioral ecologist Jeff Stevens compared the discounting behavior of two closely related New World monkeys—cotton-top tamarins and common marmosets. Both species are cooperative breeders with one dominant breeding pair and their offspring who often stick around to help rear the next generation. Both species have similar brain-size-to-body-size ratios, group sizes, and life spans. And both species live in the upper rain-forest canopies, foraging for fruits, insects, and tree sap. There are, however, two key differences: feeding specializations and territory size. Tamarins specialize in insects, whereas marmosets specialize in sap, and tamarins have significantly larger territories. These differences generate two interesting predictions. Given the tamarins' preference for insects, they should be more impulsive or impatient than marmosets, who specialize in sap. When insects are about, there is no time to wait. Foragers must attack whatever they see, immediately. In contrast,

sap feeding requires patience. The forager must scratch at the surface of the tree until it starts the flow of sap, and then sit and wait for it to ooze out; leaving the area and then coming back isn't an option, because another individual can readily profit from the original forager's efforts to break through. The differences in territory size lead to a different prediction. Given the larger size of tamarin territories, they should be willing to travel greater distances for food than marmosets. Thus, if we imagine distance as a proxy to time and effort, marmosets should devalue distant rewards more steeply than tamarins. Marmosets should settle for a small piece of food that is nearby over a large piece of food far away, whereas tamarins should be willing to travel the extra mile.

When these species worked against the clock, tamarins were impulsive and marmosets patient: Tamarins waited about half as long for the larger reward as the marmosets did. When these species worked against the tape measure, tamarins traveled significantly longer for the larger reward. Together, these results show that in our attempt to understand the evolution of patience, we must not ignore the essential role that a species' ecology plays in shaping their minds. What appears to be irrational may actually be a perfectly rational and adaptive solution under realistic constraints.

Animals are capable of extreme patience in highly specialized contexts: stashing food in birds and rodents, sit-and-wait-predators waiting for prey, and in some primates when extracting food. But in parallel with our discussion of mind reading, this is likely another example of laser-beam intelligence, a unique specialization locked into one or a few contexts, with no evidence of flexibility. What we have yet to explore, however, is whether the impatience animals show in the context of foraging extends to social situations involving violence and cooperation, problems that hook us back to morality.

DOMESTICATING VIOLENCE

Dominance hierarchies, unwritten rules of territoriality, and property ownership work well, most of the time, to control aggression. Physical aggression, harassment, and withholding resources also function in the

service of unwritten rules of punishment. These policing mechanisms are, however, weak, bound to a narrow range of contexts, and rarely if ever used in the service of moderating cooperative relationships among animals. If a lion lags behind in a context requiring cooperation, there are no costs to the laggard. If a capuchin monkey fails to help a group member acquire food, it is not beaten for its apathy. If a dolphin fails to join in on an alliance, it is not chased out to another ocean or excluded from further alliances. In the social domain, there is always someone breaking through the lines of cooperation, defecting when it pays and the costs are small. Often it is the strong over the weak, and the smart and savvy over the dolts. But the weak and dim fight back, fueling an arms race of competition.

Among animals, killing is relatively rare. Animals threaten and fight one another, but rarely attack to kill. Our own species counts as an exception, but not the only exception. The lack of killing raises two interesting questions concerning the nature of violence in animals: What stops and starts it? Are there principles of harm that guide violence in animals, paralleling some of the principles uncovered for humans? To answer the second question, we need some answers to the first, focused on what controls the impulse to fight and sometimes kill others, which we will discuss further in chapter 7.

The ethologist and Nobel laureate Konrad Lorenz suggested that the aggressive instinct is often controlled or suppressed by the submissive gestures of other individuals—a point I raised in discussing James Blair's theory of morality as viewed through the eyes of a psychopath. A snarling dog is likely to go no further if it sees its victim look away with its tail between its legs. It may refrain from an aggressive attack, because once the submissive signal is launched, there is no additional benefit from pushing further. Some authors have argued that submissive gestures work by tapping the aggressor's compassion or empathy. Empathy—feeling what another is feeling—can operate at a strictly physiological level, without any awareness. The snarling dog may stop because it feels what the subordinate feels following an attack, and this suppresses any further aggression. Empathy can also operate with awareness. Perhaps the snarling dog imagines what it would be like to be in the subordinate's place, and this turns off his aggression. At this point, there is no evidence that dogs imagine

what it is like to be another dog, but there is also no evidence to rule out this possibility.

One context associated with both aggression and conflict concerns emigration and immigration into a new group. For emigrants and residents, there are impulses that are likely to push in one direction or another: to leave or stay, to fight or flee. In socially living mammals, including most of the nonhuman primates, a tension arises in the life of a young male when he reaches reproductive maturity. He can either stay in his own natal group or leave to join another. While living with his natal group, he will have antagonistic relationships with his neighbors. But once he decides to leave, he never looks back. There are, however, costs associated with leaving. An attempt to find a suitable group with mates and an opportunity to climb up the social hierarchy will undoubtedly involve at least one good fight. From the resident's perspective, seeing a foreigner elicits curiosity, but it may also elicit aggression and a bit of fear, especially in cases where the immigrant males go on an infanticidal rampage, killing all of the new infants in the group.

Studies of wild and captive monkeys show that the hormone serotonin plays a role in these social contexts, as it does in parallel human contexts.[35] Animals with low levels of serotonin are more impulsive, emigrating from their natal group at an earlier age and approaching threatening intruders more quickly than individuals with high levels of serotonin. Adolescent males in general have lower levels of serotonin and are more impulsive, with the same pattern holding for subordinate as compared with dominant males. One can even show that serotonin causally influences social impulsivity, as opposed to being merely correlated with it. Treating animals such as vervet monkeys with the drug fluoxetine—Prozac in common parlance—decreases the uptake of serotonin, thereby increasing the levels of serotonin. Vervets with higher serotonin levels are less likely to approach a threatening intruder.

An important link to the work on serotonin and impulsivity are studies of aggression and the hormone testosterone. As Dave Barry has often mused, especially in his *Guide to Guys*, much of the chest-puffing machismo of men is due to testosterone poisoning. Fortunately, serotonin and testosterone are engaged in a physiological ballet. Testosterone motivates aggression, while serotonin regulates the level or intensity of aggression. If

testosterone levels are high, then the odds of a fight are high as well. Serotonin may then act to reduce the chances of a fight by diminishing the tendency to strike out at the slightest provocation. When serotonin levels are low, impulsivity is high, and the brain relinquishes control of aggression.

In a study of wild rhesus monkeys living on an island off the coast of South Carolina,[36] young males with high levels of testosterone frequently threatened other males, though they did not necessarily suffer any injuries. Individuals with low levels of serotonin, however, had not only more fights, but more severe injuries than individuals with high levels of serotonin. Young males with low levels of serotonin were also more likely to take leaps across large gaps in the canopy, a dangerous move that suggests risk-taking in contexts other than aggressive ones. Testosterone is trouble, as the biologist Robert Sapolsky[37] has pointed out, and what makes guys act macho—and stupid—as Dave Barry points out. Fortunately, for some animals at least, serotonin saves the day, turning knee-jerk, impulsive aggression into more controlled and calculated attacks when fighting is necessary.

Although there is ample evidence that natural selection has played a role in the design of aggressive impulses, little is known about how such selection works on the brain, how rapidly it can alter brain structure and chemistry, and the extent to which it, as opposed to other factors, has contributed to each species' aggressive profile. But there is a different way into this problem: artificial selection by means of domestication.

Anyone familiar with dog breeds will attest that there is a continuum of types running from the let-me-at-your-jugular pit bull to the please-rub-my-belly Labrador. Breeders have created this variation. For domestication to work, however, animals must lose both their fear of humans and their tendencies to be aggressive to each other. But in creating differences between breeds, and by selecting against aggressive impulses, the selective process has resulted in a series of unexpected characteristics that provide a window into the mechanisms of control.[38] Looking across domesticated animals as a group, including dogs, cats, and many farm animals, not only has there been a general reduction in aggression relative to the wild type—compare dogs with wolves, or cats with lions—but there has been an overall decrease in brain and canine size, along with an increase in what appear to be unrelated bits of anatomy, such as lop ears and coats of

fur with distinctive white splotches. All of these changes suggest that domestication leads to a shift back to juvenile qualities, or what biologists refer to as "paedomorphosis."

The most detailed study of the domestication process comes from work on the silver fox by the biologist Dmitry Belyaev. His goal was to explore the process of domestication by selecting for tameness. The technique was simple: approach a wild fox and note the distance at which it runs away. Define tameness as approach distance. Take those foxes with the shortest approach distances and breed them. Take the next generation of offspring and repeat this process. After forty years and thirty generations of artificial selection, Belyaev had produced a population of tame foxes, with newborn kits as friendly as newborn puppies. Further, and paralleling all other cases of domesticated animals, this new generation of foxes looked different, exhibiting a white patch of fur on the head, a curled tail, lop ears, and a significant reduction in skull size relative to the wild type. At a more microscopic level of the brain, the domesticated foxes also showed a higher level of serotonin. Recall that higher serotonin levels are associated with greater control over impulsivity and, thus, lower levels of knee-jerk aggression. As Belyaev reported, although they had only selected for tameness, they ended up with more than they expected: a fox with a different appearance, brain, temperament, and social savviness that comes from hanging out with humans.

The punch line, one that we can derive from hindsight, is that when humans select for a particular trait, there are always unanticipated consequences because of hidden relationships or correlations between traits. Further, although the focus and intensity of artificial selection may be different from natural selection, it is clear that selection can rapidly transform the brain of a mammal as complex as the silver fox, leading to dramatic changes in behavior. Selection can change the dynamics of the arms race, favoring either impulsivity or control.

Can we be certain that Belyaev selected for tameness, and only this characteristic? Although he used approach distance to characterize each generation, it is possible that he inadvertently selected for something else. For example, perhaps those foxes that allow humans to approach closest have higher levels of serotonin. In breeding these individuals, selection is operating on serotonin levels. Alternatively, perhaps those individuals

with the shortest approach distances are the ones that maintain eye contact, and are thus more socially skilled and attentive. Breeding these individuals would select for differences in attention or social cognition. The point of these challenges is not to undermine the results but to question their cause. Tameness is simply a description of behavior. In selecting for an outcome that we describe as tame, we don't necessarily capture the psychology that enables such behavior. A wild fox that doesn't run away from a human may do so for a variety of reasons. The fox experiments show that artificial selection can change impulsivity—over a short period of time—but they don't show how the process occurred. In terms of our moral faculty, these studies show that intense selection can rapidly change the temperament and social savviness of a complex vertebrate. This sets up a significant challenge to those who believe that the human mind was largely sculpted in the Pleio-Pleistocene period of evolution, and kept relatively mummified since. Though it is possible that we have held on to many of our hunter-gatherer thoughts and emotions, as these were surely good tricks for survival, the story of the silver fox opens the possibility of significant and rapid changes in brain evolution.

TEMPTED BY THE TRUTH
OF ANOTHER

When is it permissible for one animal to harm another? The discussion thus far suggests that animals harm others during predation, while attacking members of a neighboring group, while beating up a lower-ranking group member, during an infanticidal run, and while redirecting aggression as a mechanism to reduce postconflict tension. Paralleling our discussion of human violence, there is no single deontological principle guiding animal violence that dictates, plain and simple, that harming another is forbidden. Nor is there a principle that states that harm is permissible whenever it feels right. We explain variation in the expression of harm by appealing to principles and parameters that are grounded in action, and especially the causes and consequences of different actions. But in addition to the Rawlsian contribution, there is also a Humean component. Let's return to an earlier example to see how this might work.

When an aggressor reconciles with its victim, there is some sense in which this interaction looks like a sympathetic or perhaps empathetic response. In chapter 4, I discussed some of the work on human empathy, inspired by Hoffman's pioneering research, and mapped out more recently in terms of development and neural correlates by Nancy Eisenberg, Andrew Meltzoff, and Tania Singer. For some, empathy entails more than feeling the same way as someone else. It entails knowing or being aware of what it is like to be someone else. In its simplest formulation, empathy grows out of a mirror neuron-like system, where my perception of an event is mirrored by my enactment of the very same event. Once in place, however, this form of empathy is transformed—either in evolution or in development—by the acquisition of mind-reading skills. With this new capacity, individuals can think about what someone else feels, imagine how they would feel in the same situation, work out what would make them feel better, and from this deduce how to make the other person feel better.

Do animals have anything like the first or second form of empathy? In my discussion of empathy in humans, I mentioned the interesting observation that people who are more empathetic are more susceptible to yawning. Yawning is generally contagious. But it is really contagious if you have a big heart, unable to turn off your compassion for others. Based on this correlation between yawning and empathy, the psychologist James Anderson wondered whether other animals might also be susceptible to contagious yawning.[39] Captive chimpanzees watched videos of other chimpanzees yawning and doing other things. Though inconsistent across individuals, some individuals consistently yawned back. We can't say that the yawners are empathetic while the non-yawners are not. What we can say is that given the observation that contagious yawning is a signature of empathy in humans, it is possible that the same holds true for chimpanzees and other species. This possibility, as well as other observations of caring in animals, sets up a more specific look for empathy.

In nature, rats forage in the company of other rats and often learn from them. In the laboratory, naïve animals learn what to eat either by following knowledgeable individuals or by smelling their breath. Although rats are social eaters, they do not naturally forfeit the opportunity

---

to eat so that someone else might have a chance.[40] To examine whether one rat might forfeit the opportunity to eat because of the benefits to another, an experimenter taught a rat to press a lever for food. The experimenter then introduced a second rat into an adjacent cage and changed the wiring of the apparatus. Now, when the rat with access to the lever pressed it, he delivered a strong shock to his neighbor. This shock had not only a direct effect on the recipient, but an indirect effect on the actor rat in control of the levers. The actor actually stopped pressing for a while and thereby forfeited access to food. In so doing, the actor incurred the cost associated with hunger while relieving the recipient of pain. This is altruism, at least in the biological sense: cost to actor, benefit to recipient. It suggests that rats can control their immediate desire for food to block an action that would cause pain to another. This looks like empathy or compassion, but simpler explanations abound. Seeing another rat in pain might be aversive. When something is aversive, animals tend to stop what they are doing. Alternatively, when the rat pressing the lever sees the other in pain, he may stop for fear of retribution.

Although these results are open to various interpretations, they provide a parallel with the discounting experiments on pigeons described earlier: At some level, the actor rat must control the temptation to eat immediately. These studies differ from the discounting experiments in that the control problem is not between some food now versus more later. The choice is between some food now versus none later because pressing causes pain to another rat. Although rats initially curtail their pressing, ultimately they go back to pressing. This makes sense, given that a failure to relax control over pressing would lead to starvation. Even though it may be wrong to shock another, and even though the rat is directly responsible for the shock, self-interest carries the moment.

In a follow-up study, an experimenter taught a group of rats to press a lever to lower a suspended block of Styrofoam to the ground; if the subject failed to press the lever, the experimenter delivered a shock. Once the rats learned to press the lever, the experimenter eliminated the shock and thereby eliminated lever pressing; in the absence of either punishment or reward, motivation to press disappears. For half of the rats, the study continued with a Styrofoam block suspended by a harness and the lever available for pressing. For the other rats, the experimenter replaced the

Styrofoam with a live rat suspended by a harness, a stressful position that leads to wriggling and squealing. Rats confronted with a suspended Styrofoam block do nothing at all. Rats looking at a suspended rat immediately start pressing the lever. Although the experimenter had no intention of shocking these rats for apathy, nor rewarding them with food for pressing, they nonetheless pressed the lever and thereby lowered their compatriots, relieving them of the stress associated with suspension. This is altruism. The actor rat incurs the cost of pressing and thereby benefits the suspended individual by lowering him to safety.

What do these results tell us about the evolution of altruism and morality more generally? Perhaps seeing another in distress triggers in the actors an emotional response that blocks off the desire for more food. In many of us, seeing an elderly person struggle to open a door or carry a bag triggers an almost reflexive and sympathetic response that results in our attempt to help, as opposed to resuming lunch or a conversation. There is no control problem, because there are no alternative choices. Seeing another rat in pain or distress is sufficient to cause a sympathetic response. Alternatively, perhaps seeing another in distress is aversive. When rats experience something that is aversive, they do what they can to stop it. Pressing the lever isn't altruistic at all.

Each of these studies looks at what rats do in situations where they can help. They leave open what rats might perceive if they watched others, some acting altruistically, others selfishly. Would they prefer to interact with altruists? Would they reject the selfish individuals from joining their group? There are no answers to these questions. Until we have them, we can't distinguish between an animal's judgments or perceptions of action and their decision to act. Studies of nonhuman primates are no better off, but do move deeper into the nature of the phenomenon.

An experimenter trained a rhesus monkey to pull one of two chains in order to obtain its daily ration of food. Subjects readily complied and fed themselves. Next, the experimenter introduced another rhesus monkey into the adjacent cage and, in parallel with the rat studies, hooked up one of the chains to a machine that would deliver a shock to the newly introduced neighbor. Mirroring the rats' behavior, rhesus also stopped pulling the chains. But unlike rats, most of the rhesus showed

far greater restraint, far greater inhibitory control. Some individuals stopped pulling for five to twelve days, functionally starving themselves. The extent to which rhesus refrained from pulling was related to two important factors: experience with shock and identity of the shockee. Individuals refrained from pulling for longer periods of time if they had the experience of being shocked, if they were paired with a familiar group member as opposed to an unfamiliar member of another group, and if they were paired with another rhesus monkey, as distinct from a rabbit.

The rhesus experiments are open to the same alternative explanations as are the rat experiments. Though rhesus may feel compassion or empathy toward another in pain, they may also see the expression of pain as aversive. Seeing another in pain is aversive. Seeing a familiar cage mate in pain is more aversive than seeing an unfamiliar rhesus. Seeing a rabbit in pain is irrelevant. Rhesus may also think that all bad deeds are punished, and thus expect retaliation if they continue to eat, thereby shocking their neighbor. But even if rhesus know that pulling leads to pain, there is no reason to conclude that they stop pulling in order to alleviate another's pain. They may stop because it is distracting, or because they expect shock themselves. Although these experiments, and those on rats, do not yield clear interpretations, they raise the possibility—discussed in the next chapter—that recognition of another's emotional state may trigger an inhibitory response. As the psychologists Stephanie Preston and Frans de Waal[41] have discussed, this could happen in a completely unconscious way, recruiting circuitry in the brain that has been designed to unify how individuals act with how they perceive others acting.

In this chapter, I have pressed on the possibility that some of the core capacities underlying our moral faculty are present in nonhuman animals. We have seen that animals experience emotions that motivate morally relevant actions, including helping and harming others, as well as reconciling differences in the service of achieving some modicum of peace. We have also seen that animals are endowed with several, if not all of the core principles of action that underlie the human infants' initial state, and that these principles ultimately hook into a capacity for mind reading and some capacity for self-reflection. Differences between humans and other

animals emerge as well. Birds and mammals are remarkably impulsive, exhibiting little control in the face of temptation; their discounting curves are steep, creating problems when it comes to delaying gratification in the context of helping another at a personal cost. Perhaps the most intriguing difference is that whereas individual species exhibit some subset of these capacities, only humans appear to have evolved a complete set.

7

FIRST PRINCIPLES

We can imagine a society in which no one could survive as a social being because it does not correspond to biologically determined perceptions and human social needs. For historical reasons, existing societies might have such properties, leading to various forms of pathology.
—NOAM CHOMSKY[1]

THROUGHOUT HISTORY, and in all the world's cultures, various groups have articulated various versions of the Golden Rule. Sometimes it has been stated with a positive angle, sometimes a negative one. The general principle has, however, always been the same:[2]

BUDDHISM: "Hurt not others in ways that you yourself would find hurtful."

CONFUCIANISM: "Surely it is a maxim of loving kindness: Do not unto others what you would not have them do unto you."

TAOISM: "Regard your neighbor's gain as your own gain and your neighbor's loss as your own loss."

JUDAISM: "What is hateful to you, do not to your fellow men. That is the entire Law; all the rest is commentary."

CHRISTIANITY: "All things whatsoever ye would that men should do to you, do ye even so to them; for this is the Law and the Prophets."