Morality and the Social Instincts: Continuity with the Other Primates

FRANS B. M. DE WAAL

THE TANNER LECTURES ON HUMAN VALUES

Delivered at

Princeton University
November 19–20, 2003

FRANS B. M. DE WAAL is C. H. Candler Professor of Primate Behavior at Emory University and director of the Living Links Center at the Yerkes Regional Primate Research Center. He was educated at three Dutch universities—Nijmegen, Groningen, and Utrecht—and received a Ph.D. from the University of Utrecht. He has conducted research on the world's largest captive colony of chimpanzees at the Arnhem Zoo, did both observational and experimental studies of reconciliation behavior in monkeys at the Wisconsin Regional Primate Research Center, and worked with bonobos at the San Diego Zoo. He is a foreign associate of the National Academy of Sciences. In addition to many scientific papers, he is the author of *Chimpanzee Politics: Power and Sex among Apes* (1982); *Peacemaking among Primates* (1989), which was awarded the *Los Angeles Times* Book Award; *Bonobo: The Forgotten Ape* (1997); *The Ape and the Sushi Master* (2001); *Tree of Origin: What Primate Behavior Tells Us about Human Social Evolution* (2001), and *My Family Album: Thirty Years of Primate Photography* (2003).

Abstract

The *Homo homini lupus* view of our species is recognizable in an influential school of biology, founded by Thomas Henry Huxley, which holds that we are born nasty and selfish. According to this school, it is only with the greatest effort that we can hope to become moral. This view of human nature is discussed here as "Veneer Theory," meaning that it sees morality as a thin layer barely disguising less noble tendencies. Veneer Theory is contrasted with the idea of Charles Darwin that morality is a natural outgrowth of the social instincts, hence continuous with the sociality of other animals.

Veneer Theory is criticized at two levels. First, it suffers from major unanswered theoretical questions. If true, we would need to explain why humans, and humans alone, have broken with their own biology, how such a feat is at all possible, and what motivates humans all over the world to do so. The Darwinian view, in contrast, has seen a steady stream of theoretical advances since the 1960s, developed out of the theories of kin selection and reciprocal altruism, but now reaching into fairness principles, reputation building, and punishment strategies. Second, Veneer Theory remains unsupported by empirical evidence. Given that it views morality as a recent addition to human behavior, it would predict that morality resides entirely in the newest parts of our enlarged brain and leads to behavior that deviates from anything other animals do. Modern neuroscience, however, has demonstrated that ethical dilemmas activate ancient emotional centers in the brain that originated long before our species. Moreover, studies of nonhuman primates hint at continuity in many areas considered relevant for an evolved morality. Human moral decisions often stem from emotionally driven "gut" reactions, some of which we share with our closest relatives. These animals may not be moral beings, but they do show signs of empathy, reciprocity, a sense of fairness, and social regularities that—like the norms and rules governing human moral conduct—promote a mutually satisfactory *modus vivendi*.

We approve and we disapprove because we cannot do otherwise.
Can we help feeling pain when the fire burns us?
Can we help sympathizing with our friends?

Edward Westermarck (1912 [1908]: 19)

Why should our nastiness be the baggage of an apish past and our kindness uniquely human? Why should we not seek continuity with other animals for our "noble" traits as well?

Stephen Jay Gould (1980: 261)

Homo homini lupus—"man is a wolf to man"—is an old Roman proverb popularized by Thomas Hobbes. Even though it permeates large parts of law, economics, and political science, the proverb contains two major errors. First, it fails to do justice to canids, which are among the most gregarious and cooperative animals on the planet (Lorenz 1954; Schleidt and Shalter 2003). But even worse, the saying denies the inherently social nature of our own species.

Social contract theory, and Western civilization with it, seems saturated with the assumption that we are asocial, even nasty creatures rather than the *zoon politikon* (political animal) that Aristotle saw in us. Hobbes explicitly rejected the Aristotelian view by proposing that our ancestors started out autonomous and combative, establishing community life only when the cost of continual strife became unbearable. Social life did not come naturally to us: the step was taken reluctantly, or in the words of Hobbes (1991 [1651]: 120), "by covenant only, which is artificial." More recently, John Rawls (1972) has proposed a milder version of this view, adding that humanity's step toward sociality hinged on conditions of fairness, that is, the prospect of mutually advantageous cooperation among equals.

These ideas about the origin of the well-ordered society remain popular even though the underlying assumption of a rational decision by inherently asocial creatures is untenable in light of what we know about the evolutionary background of our species. It creates the illusion of human society as a voluntary arrangement with self-imposed rules assented to by free and equal persons. Yet there never was a point at which we became social: descended from highly social ancestors, the monkeys and apes, we have been group-living forever. Free and equal

people never existed. Humans started out—if a starting point is discernible at all—as interdependent, bonded, and unequal. We come from a long lineage of hierarchical animals for which life in groups is not an option but a survival strategy. Having companions offers advantages in locating food and avoiding predators (Wrangham 1980; van Schaik 1983). Inasmuch as group-oriented, gregarious individuals left more offspring than those less socially inclined (e.g., Silk et al. 2003), sociality became ever deeper ingrained in primate biology and psychology. If any decision to establish societies was made at all, therefore, credit should go to Mother Nature instead of ourselves.

This is not to dismiss the heuristic value of Rawls's "original position" as a way of getting us to reflect on what kind of society we would *prefer* to live in. The original position refers to a "purely hypothetical situation characterized so as to lead to certain conceptions of justice" (Rawls 1972: 12). But even if we do not take the original position literally, adopting it only for the sake of argument, it still distracts from the more pertinent argument that we should be pursuing instead about how we actually became what we are today. What parts of human nature have led us down this or that path, and how have these parts been shaped by evolution? Addressing a real rather than hypothetical past, such questions are bound to bring us closer to the truth. The truth is that we are born intensely social.

A good illustration of the social nature of our species is that, second to the death penalty, solitary confinement is the most extreme punishment we can think of. It works this way only, of course, because we are not born as loners. Our bodies and minds are not designed for life without others. We become hopelessly depressed in the absence of company. Without social support, our health deteriorates. In one recent experiment, healthy volunteers deliberately exposed to cold and flu viruses got sick more easily if they had fewer friends and family around (Cohen et al. 1997). While the primacy of connectedness is naturally understood by women—perhaps because mammalian females with caring tendencies have outreproduced those without for 180 million years—it applies equally to men. In modern society, there is no more effective way for men to expand their age horizon than to get and stay married: it increases their chance of living past the age of sixty-five from 65 to 90 percent (Taylor 2002).

Our social makeup is so obvious that there would be no need to belabor this point were it not for its conspicuous absence from origin stories within the disciplines of law, economics, and political science. A tendency

in the West to see emotions as soft and social attachment as messy has made theoreticians turn to cognition and rationality as the preferred guides of human behavior. This is so despite the fact that psychological research suggests the primacy of affect: that is, that human behavior derives above all from fast, automated emotional judgments and only secondarily from slower conscious processes (e.g., Zajonc 1980, 1984; Bargh and Chartrand 1999). Humans seem, in fact, about as emotional in their dealing with each other as any other social animal.

Unfortunately, the overemphasis on rationality and downplaying of emotions is not restricted to the humanities and social sciences. Within evolutionary biology, too, some have embraced the illusion that we are a self-invented species. A parallel debate pitting reason against emotion has been raging regarding the origin of morality, a hallmark of human society. One school views morality as a cultural innovation achieved by our species alone. This school does not see moral tendencies as part and parcel of human nature. Our ancestors, it claims, became moral by choice. The second school, in contrast, views morality as growing out of the social instincts that we share with many other animals. In this view, morality is neither unique to us nor a conscious decision taken at a specific point in time: it is the product of gradual social evolution.

The first standpoint assumes that deep down we are not truly moral. It views morality as a cultural overlay, a thin veneer hiding an otherwise selfish and brutish nature. Perfectibility is what we should strive for. Until recently, this was the dominant view within evolutionary biology as well as among science writers popularizing this field. I use the term "Veneer Theory" to denote these ideas, tracing their origin to Thomas Henry Huxley. After treating these ideas, I review Charles Darwin's quite different standpoint of an evolved morality, which was inspired by the Scottish Enlightenment. I further discuss the views of Mencius and Edward Westermarck, which agree with those of Darwin.

Given these two schools' contrasting opinions about continuity versus discontinuity with other animals, I build upon an earlier treatise (de Waal 1996) in paying special attention to parallels between the behavior of human and nonhuman primates.

1. Veneer Theory

In 1893, for a large audience in Oxford, England, Huxley publicly reconciled his dim view of the natural world with the kindness occasionally encountered in human society. Huxley realized that the laws of the phys-

ical world are unalterable. He felt, however, that their impact on human existence could be softened and modified if people kept nature under control. Huxley compared us with a gardener who has a hard time keeping weeds out of his garden. He declared human ethics a cultural victory over the evolutionary process (Huxley 1989 [1894]).

This was an astounding position for two reasons. First, it deliberately curbed the explanatory power of evolution. Since many consider morality the essence of our species, Huxley was in effect saying that what makes us human could not be handled by the evolutionary framework. This was an inexplicable retreat by someone who had gained a reputation as "Darwin's Bulldog" owing to his fierce advocacy of evolutionary theory. Second, Huxley gave no hint whatsoever where humanity might have unearthed the will and strength to go against its own nature. If we are indeed born competitors, who don't care about the feelings of others, how did we decide to transform ourselves into model citizens? Can people for generations maintain behavior that is out of character, like a shoal of piranhas that decides to turn vegetarian? How deep does such a change go? Are we the proverbial wolves in sheep's clothing: nice on the outside, nasty on the inside?

This was the only time Huxley visibly broke with Darwin. As Huxley's biographer, Adrian Desmond (1994: 599), put it: "Huxley was forcing his ethical Ark against the Darwinian current which had brought him so far." Two decades earlier, in *The Descent of Man,* Darwin (1982 [1871]) had unequivocally stressed morality as part of human nature. The reason for Huxley's departure has been sought in his suffering at the cruel hand of nature, which had taken the life of his beloved daughter, as well as his need to make the ruthlessness of the Darwinian cosmos palatable to the general public. He had depicted nature as so thoroughly "red in tooth and claw" that he could maintain this position only by dislodging human ethics, presenting it as a separate innovation (Desmond 1994). In short, Huxley had talked himself into a corner.

His curious dualism, which pits morality against nature and humans against other animals, was to receive a respectability boost from Sigmund Freud's writings, which throve on contrasts between the conscious and subconscious, the ego and superego, Love and Death, and so on. As with Huxley's gardener and garden, Freud was not just dividing the world in symmetrical halves: he saw struggle everywhere. He explained the incest taboo and other moral restrictions as the result of a violent break with the freewheeling sexual life of the primal horde, culminating in the

collective slaughter of an overbearing father by his sons (Freud 1913). He let civilization arise out of the renunciation of instinct, the gaining of control over the forces of nature, and the building of a cultural superego.

Man's heroic combat against forces that try to drag him down remains a dominant theme within biology today, as illustrated by quotes from outspoken Huxleyans. Declaring ethics a radical break with biology, George Williams has written extensively about the wretchedness of nature, culminating in his claim that human morality is a mere by-product of the evolutionary process: "I account for morality as an accidental capability produced, in its boundless stupidity, by a biological process that is normally opposed to the expression of such a capability" (Williams 1988: 438).

Having explained at length that our genes know what is best for us, programming every little wheel of the human survival machine, Richard Dawkins waits until the last sentence of *The Selfish Gene* to reassure us that, in fact, we are welcome to chuck all of those genes out of the window: "We, alone on earth, can rebel against the tyranny of the selfish replicators" (Dawkins 1976: 215). The break with nature is obvious in this statement, as is the uniqueness of our species. Dawkins explicitly endorses Huxley: "What I am saying, along with many other people, among them T. H. Huxley, is that in our political and social life we are entitled to throw out Darwinism, to say we don't want to live in a Darwinian world" (Roes 1997: 3; see also Dawkins 2003).

Darwin must be turning in his grave, because the implied "Darwinian world" is miles removed from what he himself envisioned (see below). What is lacking in these statements is any indication of how we can possibly negate our genes, which the same authors at other times don't hesitate to depict as all-powerful. Like the views of Hobbes, Huxley, and Freud, the thinking is thoroughly dichotomous: we are part nature, part culture, rather than a well-integrated whole. Morality is a thin crust underneath which boil human passions that are invariably antisocial, amoral, and egoistic. This idea of morality as a veneer is best summarized by Michael Ghiselin's famous quip: "Scratch an 'altruist,' and watch a 'hypocrite' bleed" (Ghiselin 1974: 247; figure 1).

Veneer Theory has been popularized by countless science writers, such as Robert Wright (1994), who went so far as to claim that virtue is absent from people's hearts and souls, that our species is potentially but not naturally moral. But what, one might ask, about the many people

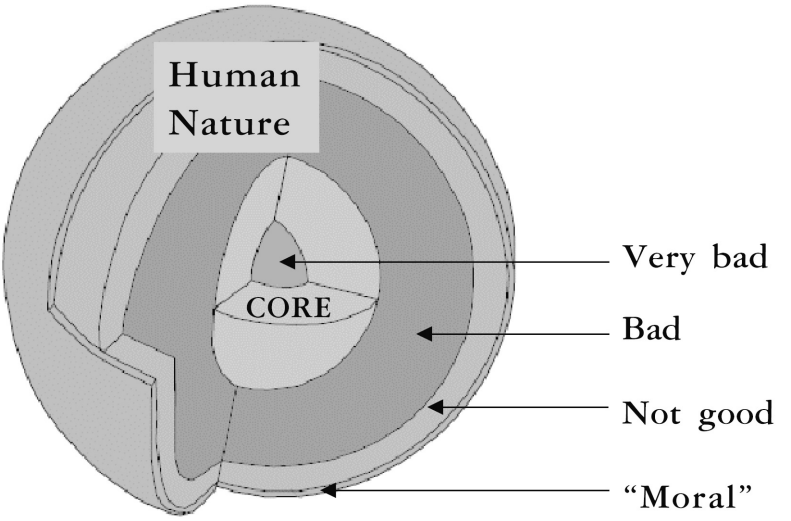

FIGURE 1. The popular view of morality among biologists during the past quarter of a century was best summarized by Ghiselin (1974: 247): "Scratch an 'altruist,' and watch a 'hypocrite' bleed." Accordingly, people are thoroughly competitive, and morality is no more than a last-minute, artificial addition. Summarized as "Veneer Theory," this idea goes back not to Charles Darwin but to his contemporary Thomas Henry Huxley. It is visualized here tongue-in-cheek as a human nature that is bad and selfish to its core.

who occasionally experience in themselves and others a degree of sympathy, goodness, and generosity? Echoing Ghiselin, Wright answers that the "moral animal" is essentially a hypocrite:

> . . .the pretense of selflessness is about as much part of human nature as is its frequent absence. We dress ourselves up in tony moral language, denying base motives and stressing our at least minimal consideration for the greater good; and we fiercely and self-righteously decry selfishness in others. (Wright 1994: 344)

To explain how we manage to live with ourselves despite this travesty, theorists have called upon self-deception. If people think they are at times unselfish, so the argument goes, they must be hiding their true motives from themselves (e.g., Badcock 1986). In the ultimate twist of irony, anyone who fails to believe that we are fooling ourselves, and feels that genuine kindness actually exists in the world, is considered a wish-

ful thinker, hence accused of fooling him- or herself. Some scientists have objected, however:

> It is frequently said that people endorse such hypotheses [about human altruism] because they *want* the world to be a friendly and hospitable place. The defenders of egoism and individualism who advance this criticism thereby pay themselves a compliment; they pat themselves on the back for staring reality squarely in the face. Egoists and individualists are objective, they suggest, whereas proponents of altruism and group selection are trapped by a comforting illusion. (Sober and Wilson 1998: 8–9)

All of these back-and-forth arguments about how to reconcile everyday human kindness with evolutionary theory seem an unfortunate legacy of Huxley, who had a poor understanding of the theory that he so zealously defended (Mayr 1997). It should be pointed out that in Huxley's time there was already opposition to his ideas (Desmond 1994), some of which came from Russian biologists, such as Petr Kropotkin. Given the harsh climate of Siberia, Russian scientists traditionally were far more impressed by the battle of animals against the elements than against each other, resulting in an emphasis on cooperation and solidarity that contrasted with Huxley's dog-eat-dog perspective (Todes 1989). Kropotkin's (1972 [1902]) *Mutual Aid* was a direct attack on Huxley, but written with great deference for Darwin.

Although Kropotkin never formulated his theory with the precision and evolutionary logic available to Robert Trivers (1971) in his seminal paper on reciprocal altruism, both pondered the origins of a cooperative, and ultimately moral, society without invoking false pretense, Freudian denial schemes, or cultural indoctrination. In this they proved the true followers of Darwin.

2. Darwin on the Evolution of Ethics

Evolution favors animals that assist each other if by doing so they achieve long-term benefits of greater value than the benefits derived from going it alone and competing with others. Unlike cooperation resting on simultaneous benefits to all parties involved (known as mutualism), reciprocity involves exchanged acts that, while beneficial to the recipient, are costly to the performer (Dugatkin 1997). This cost, which is generated because there is a time lag between giving and receiving, is

eliminated as soon as a favor of equal value is returned to the performer (for treatments of this issue since Trivers 1971, see Axelrod and Hamilton 1981; Rothstein and Pierotti 1988; Taylor and McGuire 1988). It is in these theories that we find the germ of an evolutionary explanation that escaped Huxley.

It is important to clarify that these theories do not conflict by any means with popular ideas about the role of selfishness in evolution. It is only recently that the concept of "selfishness" has been plucked from the English language, robbed of its vernacular meaning, and applied outside of the psychological domain. Even though the term is seen by some as synonymous with being self-serving, English does have different terms for a reason. Selfishness implies the *intention* to serve oneself, hence knowledge of what one stands to gain from a particular behavior. A vine may be self-serving by overgrowing and suffocating a tree; but since plants lack intentions, they cannot be selfish except in a meaningless, metaphorical sense. Unfortunately, in complete violation of the term's original meaning, it is precisely this rather empty sense of "selfish" that has come to dominate debates about human nature. If our genes are selfish, we must be selfish, too, is the argument one often hears, despite the fact that genes are mere molecules, and hence cannot be selfish (Midgley 1979).

It is fine to describe animals (and humans) as the product of evolutionary forces that promote self-serving behavior so long as one realizes that this by no means precludes the evolution of altruistic and sympathetic tendencies. Darwin recognized this, explaining the evolution of these tendencies by group selection instead of the individual and kin selection favored by modern theoreticians (but see, e.g., Sober and Wilson 1998; Boehm 1999). Darwin firmly believed his theory capable of accommodating the origins of morality and did not see any conflict between the harshness of the evolutionary process and the gentleness of some of its products. Rather than presenting the human species as outside of the laws of biology, Darwin emphasized continuity with animals even in the moral domain:

> 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 1982 [1871]: 71–72)

It is important to dwell on the capacity for sympathy hinted at here and expressed more clearly by Darwin elsewhere (e.g., "Many animals certainly sympathize with each other's distress or danger" [Darwin 1982 (1871): 77]), because it is in this domain that striking continuities exist between humans and other social animals. To be vicariously affected by the emotions of others must be very basic, because these reactions have been reported for a great variety of animals and are often immediate and uncontrollable. They undoubtedly derive from parental care, in which vulnerable individuals are fed and protected, but in many animals stretch well beyond this domain, extending to relations among unrelated adults (section 4 below).

In his view of sympathy, Darwin was inspired by Adam Smith, the Scottish moral philosopher and father of economics. It says a great deal about the distinctions we need to make between self-serving behavior and selfish motives that Smith, best known for his emphasis on self-interest as the guiding principle of economics, also wrote about the universal human capacity of sympathy:

> How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it. (Smith 1937 [1759]: 9)

The evolutionary origin of this inclination is no mystery. All species that rely on cooperation—from elephants to wolves and people—show group loyalty and helping tendencies. These tendencies evolved in the context of a close-knit social life in which they benefited relatives and companions able to repay the favor. The impulse to help was therefore never totally without survival value to the ones showing the impulse. But, as so often, the impulse became divorced from the consequences that shaped its evolution. This permitted its expression even when payoffs were unlikely, such as when strangers were beneficiaries. Personally, I am unconvinced that we need group selection to explain the origin of these tendencies—we seem to be able to get quite far with the theories of kin selection and reciprocal altruism. Moreover, there is so much intergroup migration (hence gene-flow) in primates that the conditions for group selection do not seem to be fulfilled. In all of the primates, the younger generation of one sex or another (males in many monkeys, females in the case of chimpanzees and bonobos) tends to leave the group

to join neighboring groups (Pusey and Packer 1987). This means that primate groups are far from genetically isolated.

In discussing what constitutes morality, the actual behavior is less important than the underlying capacities. For example, instead of arguing that food-sharing is a building block of morality, it is rather the capacities thought to underlie food-sharing (e.g., high levels of tolerance, sensitivity to others' needs, reciprocal exchange) that are relevant. Ants, too, share food, but likely based on quite different urges than those that make chimpanzees or people share food (de Waal 1989a). This distinction was understood by Darwin, who looked beyond the actual behavior at the underlying motivations, intentions, and capacities. In other words, whether animals are nice to each other is not the issue; nor does it matter much whether their behavior fits our moral preferences or not. The relevant question is whether they possess the capacities for reciprocity and revenge, for the enforcement of social rules, for the settlement of disputes, and for sympathy and empathy (Flack and de Waal 2000).

This also means that calls to reject Darwinism in our daily lives so as to build a moral society are based on a misreading of Darwin. Since Darwin saw morality as a logical evolutionary product, he envisioned an eminently more livable world than the one proposed by Huxley and his followers. The latter believe in a culturally imposed, artificial morality that seems impossible to maintain, given that human nature is offering no helping hand. Huxley's world seems by far the colder and more terrifying place.

3. Edward Westermarck

Edward Westermarck, a Swedish Finn who lived from 1862 until 1939, deserves a central position in any debate about the origin of morality, since he was the first scholar to promote an integrated view including both humans and animals and both culture and evolution. That his ideas were underappreciated at the time is understandable, because they flew in the face of the Western dualistic tradition that pits body against mind and culture against instinct.

Westermarck's books are a curious blend of dry theorizing, detailed anthropology, and secondhand animal stories. The author was eager to connect human and animal behavior, but his own work focused entirely on people. Since in his days little systematic research on animal behavior existed, he had to rely on anecdotes, such as the one of a vengeful camel

that had been excessively beaten on multiple occasions by a fourteen-year-old "lad" for loitering or turning the wrong way. The camel passively took the punishment; but a few days later, finding itself unladen alone on the road with the same conductor, "seized the unlucky boy's head in its monstrous mouth, and lifting him up in the air flung him down again on the earth with the upper part of the skull completely torn off, and his brains scattered on the ground" (Westermarck 1912 [1908]: 38).

We should not discard such unverified reports out of hand: stories of delayed retaliation abound in the zoo world, especially about apes and elephants. We now have systematic data on how chimpanzees punish negative actions with other negative actions (called a "revenge system" by de Waal and Luttrell 1988), and how a macaque attacked by a dominant member of its troop will turn around to redirect aggression against a vulnerable younger relative of its attacker (Aureli et al. 1992). These reactions fall under Westermarck's *retributive emotions*, but for him the term "retributive" went beyond its usual connotation of getting even. It also covered positive emotions, such as gratitude and the repayment of services. Depicting the retributive emotions as the cornerstone of morality, Westermarck weighed in on the question of its origin while antedating modern discussions of evolutionary ethics.

Westermarck is part of a long tradition, going back to Aristotle and Thomas Aquinas, which firmly anchors morality in the natural inclinations and desires of our species (Arnhart 1998, 1999). Emotions occupy a central role; it is well known that, rather than being the antithesis of rationality, emotions aid human reasoning. People can reason and deliberate as much as they want: neuroscientists have found that if there are no emotions attached to the various options in front of them, they will never reach a decision or conviction (Damasio 1994). This is critical for moral choice, because if anything morality involves strong convictions. These convictions don't—or rather can't—come about through a cool rationality: they require caring about others and powerful "gut feelings" about right and wrong.

Westermarck (1912 [1908], 1917 [1908]) discusses, one by one, a whole range of what philosophers before him, most notably David Hume (1978 [1739]) and Adam Smith (1937 [1759]), called the "moral sentiments." He classified the retributive emotions into those derived from resentment and anger, which seek revenge and punishment, and those that are more positive and prosocial. Whereas in his time few animal examples existed of the moral emotions—hence his reliance on

Moroccan camel stories—we know now that there are many parallels in primate behavior. He also discusses "forgiveness," and how the turning of the other cheek is a universally appreciated gesture. Chimpanzees kiss and embrace after fights, and these so-called reconciliations serve to preserve peace within the community (de Waal and van Roosmalen 1979). A rapidly growing literature exists on conflict resolution in primates and other mammals (de Waal 1989b, 2000; Aureli and de Waal 2000; Aureli et al. 2002). Reconciliation may not be the same as forgiveness, but the two are obviously related.

Westermarck also sees protection of others against aggression as resulting from what he calls "sympathetic resentment," thus implying that this behavior rests on identification and empathy with the other. Protection against aggression is common in monkeys and apes and in many other animals, who stick up for their kin and friends. The primate literature offers a well-investigated picture of coalitions and alliances, which some consider the hallmark of primate social life and the main reason that primates have evolved such complex, cognitively demanding societies (e.g., Byrne and Whiten 1988; Harcourt and de Waal 1992; de Waal 1998 [1982]).

Similarly, the retributive kindly emotions ("desire to give pleasure in return for pleasure": Westermarck 1912 [1908]: 93) have an obvious parallel in what we now call reciprocal altruism, such as the tendency to repay in kind those from whom assistance has been received. Westermarck adds moral approval as a retributive kindly emotion, hence as a component of reciprocal altruism. These views antedate the discussions about "indirect reciprocity" in the modern literature on evolutionary ethics, which revolve around reputation building within the larger community (e.g., Alexander 1987). It is truly amazing to see how many issues brought up by contemporary authors are, couched in somewhat different terms, already present in the writings of this Swedish Finn of one century ago.

The most insightful part of Westermarck's work is perhaps where he tries to come to grips with what defines a moral emotion as moral. Here he shows that there is more to these emotions than raw gut feeling, as he explains that they "differ from kindred non-moral emotions by their disinterestedness, apparent impartiality, and flavour of generality" (Westermarck 1917 [1908]: 738–39). Emotions, such as gratitude and resentment, directly concern one's own interests—how one has been treated or how one wishes to be treated—hence they are too egocentric

to be moral. Moral emotions ought to be disconnected from one's immediate situation: they deal with good and bad at a more abstract, disinterested level. It is only when we make general judgments of how *anyone* ought to be treated that we can begin to speak of moral approval and disapproval. It is in this specific area, famously symbolized by Smith's (1937 [1759]) "impartial spectator," that humans seem to go radically further than other primates.

Section 4 discusses continuity between the two main pillars of human morality and primate behavior. Empathy and reciprocity have been described as the chief "prerequisites" (de Waal 1996) or "building blocks" of morality (Flack and de Waal 2000), meaning that whereas they are by no means sufficient to produce morality as we know it, they are indispensable.

4. Animal Empathy

4a. Emotional Linkage

When Carolyn Zahn-Waxler visited homes to find out how children respond to family members instructed to feign sadness (sobbing), pain (crying), or distress (choking), she discovered that children a little over one year of age already comfort others. This is a milestone in their development: an aversive experience in another person draws out a concerned response. An unplanned sidebar to her classical study, however, was that household pets appeared as worried as the children by the "distress" of family members. They hovered over them or put their heads in their laps (Zahn-Waxler et al. 1984).

Intersubjectivity has many aspects apart from emotional linkage, such as an appraisal of the other's situation, experience-based predictions about the other's behavior, extraction of information from the other that is valuable to the self, and an understanding of the other's knowledge and intentions. When the emotional state of one individual induces a matching or related state in another, we speak of *emotional contagion* (Hatfield et al. 1993). With increasing differentiation between self and other, and an increasing appreciation of the precise circumstances underlying the emotional states of others, emotional contagion develops into empathy. Empathy encompasses—and could not possibly exist without—emotional contagion, yet goes beyond it in that it places filters between the other's state and the own, adding a cognitive layer. In empathy, the subject does *not* confuse its own internal state with the other's. These

various levels of empathy, including personal distress and sympathetic concern, are defined and discussed by Nancy Eisenberg (2000).

Empathy is a social phenomenon with great adaptive significance for animals in groups. That most modern textbooks on animal cognition do not index empathy or sympathy does not mean that these capacities are not essential; it only means that they have been overlooked by a science traditionally focused on individual rather than interindividual capacities. Inasmuch as the survival of many animals depends on concerted action, mutual aid, and information transfer, selection must have favored proximate mechanisms to evaluate the emotional states of others and quickly respond to them in adaptive ways. Even though the human empathy literature often emphasizes the cognitive side of this ability, proposing complex simulations or evaluations of the other's state, Martin Hoffman (1981b: 79) rightly notes that "humans must be equipped biologically to function effectively in many social situations without undue reliance on cognitive processes."

Empathy, which allows us to relate to the emotional states of others, seems critical for the regulation of social interactions, such as coordinated activity, cooperation toward a common goal, social bonding, and care of others. It would be strange indeed if such an essential survival mechanism, which arises so early in life in all members of our species, would totally lack animal parallels.

4b. Early Experiments

An interesting older literature by experimental psychologists (reviewed by Preston and de Waal 2002a and 2002b and de Waal 2003) placed the words "empathy" and "sympathy" between quotation marks. In those days, talk of animal emotions was all but taboo. In a paper provocatively entitled "Emotional Reactions of Rats to the Pain of Others," R. M. Church (1959) established that rats that had learned to press a lever to obtain food would stop doing so if their response was paired with the delivery of an electric shock to a visible neighboring rat. Even though this inhibition habituated rapidly, it suggested something aversive about the pain reactions of others. Perhaps such reactions arouse negative emotions in rats that see and hear them.

Monkeys show a stronger inhibition than rats. The most compelling evidence for the strength of empathy in monkeys came from S. Wechkin et al. (1964) and J. Masserman et al. (1964). They found that rhesus monkeys refuse to pull a chain that delivers food to themselves if doing

so shocks a companion. One monkey stopped pulling for five days, and another one for twelve days after witnessing shock-delivery to a companion. These monkeys were literally starving themselves to avoid inflicting pain upon another. Such sacrifice relates to the tight social system and emotional linkage among macaques, as supported by the finding that the inhibition to hurt another was more pronounced between familiar rather than unfamiliar individuals (Masserman et al. 1964).

4c. Consolation Behavior

Qualitative accounts of great ape temperament support the view that these animals show strong emotional reactions to others in pain or need. Thus, Robert Yerkes (1925: 246) reports how his bonobo was so extraordinarily concerned and protective toward his sickly chimpanzee companion, Panzee, that the scientific establishment might not accept his claims: "If I were to tell of his altruistic and obviously sympathetic behavior towards Panzee I should be suspected of idealizing an ape."

Nadezhda Ladygina-Kohts (2001 [1935]: 121) noticed similar empathic tendencies in her young chimpanzee, Joni, whom she raised at the beginning of the previous century in Moscow. Kohts, who analyzed Joni's behavior in the minutest detail, discovered that the only way to get him off the roof of her house after an escape (much better than any reward or threat of punishment) was by appealing to his sympathy:

> If I pretend to be crying, close my eyes and weep, Joni immediately stops his plays or any other activities, quickly runs over to me, all excited and shagged, from the most remote places in the house, such as the roof or the ceiling of his cage, from where I could not drive him down despite my persistent calls and entreaties. He hastily runs around me, as if looking for the offender; looking at my face, he tenderly takes my chin in his palm, lightly touches my face with his finger, as though trying to understand what is happening, and turns around, clenching his toes into firm fists.

These are just two out of many reports gathered and discussed by de Waal (1996, 1997a) that suggest that apart from emotional connectedness apes have an appreciation of the other's situation and a degree of perspective-taking. Another striking report in this regard concerns a bonobo female empathizing with a bird at Twycross Zoo, in England:

> One day, Kuni captured a starling. Out of fear that she might molest the stunned bird, which appeared undamaged, the keeper urged the

> ape to let it go.... Kuni picked up the starling with one hand and climbed to the highest point of the highest tree where she wrapped her legs around the trunk so that she had both hands free to hold the bird. She then carefully unfolded its wings and spread them wide open, one wing in each hand, before throwing the bird as hard she could towards the barrier of the enclosure. Unfortunately, it fell short and landed onto the bank of the moat where Kuni guarded it for a long time against a curious juvenile. (de Waal 1997a: 156)

Obviously, what Kuni did would have been totally inappropriate toward a member of her own species. Having seen birds in flight many times, she seemed to have a notion of what would be good for a bird, thus giving us an anthropoid illustration of the empathic capacity so enduringly described by Smith (1937 [1759]: 10) as "changing places in fancy with the sufferer."

Primate empathy is such a rich area that Sanjida O'Connell (1995) was able to conduct a content analysis of thousands of qualitative reports. The investigator counted the frequency of three types of empathy, from emotional contagion to more cognitive forms, including an appreciation of the other's situation and aid-giving that is tailored to the other's needs. Understanding the emotional state of another was particularly common in the chimpanzee, with most outcomes resulting in the subject comforting the object of distress. Monkey displays of empathy were far more restricted but did include the adoption of orphans and reactions to illness, handicaps, and wounded companions.

This difference between monkey and ape empathy has been confirmed by systematic studies of behavior known as "consolation," first documented by de Waal and Angeline van Roosmalen (1979). Consolation is defined as reassurance and friendly contact directed by an uninvolved bystander to one of the combatants in a preceding aggressive incident. For example, a third party goes over to the loser of a fight and gently puts an arm around her shoulders (figure 2). Consolation is not to be confused with reconciliation, which seems mostly self-interested, such as by the imperative to restore a disturbed social relationship (de Waal 2000). The advantages of consolation for the actor remain unclear. The actor could probably walk away from the scene without any negative consequences.

Information on chimpanzee consolation is well quantified. De Waal and van Roosmalen (1979) based their conclusions on an analysis of hundreds of postconflict observations, and a replication study by de Waal

FIGURE 2. A typical instance of consolation in chimpanzees in which a juvenile puts an arm around a screaming adult male who has just been defeated in a fight with his rival. Photograph by the author.

and Filippo Aureli (1996) included an even larger sample in which the authors sought to test two relatively simple predictions. If third-party contacts indeed serve to alleviate the distress of conflict participants, these contacts should be directed more at recipients of aggression than at aggressors, and more at recipients of intense rather than mild aggression. Comparing third-party contact rates with baseline levels, the investigators found support for both of these predictions (figure 3).

Consolation has thus far been demonstrated in great apes only. When de Waal and Aureli (1996) set out to apply exactly the same observation methodology as used on chimpanzees to detect consolation in macaques, they failed to find any (reviewed by Watts et al. 2000). This came as a surprise, because reconciliation studies, which employ essentially the same design, have shown reconciliation in species after species. Why, then, would consolation be restricted to apes?

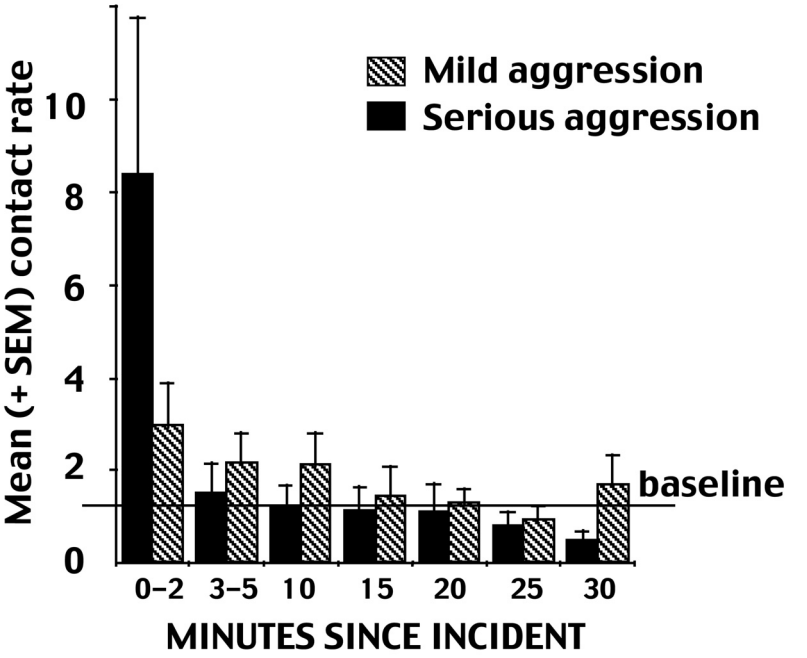

FIGURE 3. The rate with which third parties contact victims of aggression in chimpanzees, comparing recipients of serious and mild aggression. Especially in the first few minutes after the incident, recipients of serious aggression receive more contacts than baseline. After de Waal and Aureli (1996).

Targeted help in response to specific, sometimes novel, situations may require a distinction between self and other that allows the other's situation to be divorced from one's own while maintaining the emotional link that motivates behavior. Possibly, one cannot achieve cognitive empathy without a high degree of self-awareness. In other words, in order to understand that the source of vicarious arousal is not oneself but the other and to understand what caused the other's state, one needs a clear distinction between self and other. Based on these assumptions, Gordon Gallup (1982) was the first to speculate about a possible connection between cognitive empathy and mirror self-recognition (MSR). This view is supported both developmentally, by a correlation between the emergence of MSR in young children and their helping tendencies (Bischof-Köhler 1988; Zahn-Waxler et al. 1992), and phylogenetically, by the presence of complex helping and consolation in hominoids, such

as humans and apes, but not in monkeys. Hominoids are also the only primates with MSR.

I have argued before that, apart from consolation behavior, targeted helping reflects cognitive empathy. Targeted helping is defined as altruistic behavior tailored to the specific needs of the other in novel situations, such as the previously described reaction of Kuni to the bird or the highly publicized case of Binti-Jua, a gorilla female who rescued a boy who had fallen into her enclosure at the Brookfield Zoo in Chicago (de Waal 1996, 2001). These responses require an understanding of the specific predicament of the individual needing help. Targeted helping is common in the great apes, but also striking in dolphins (Caldwell and Caldwell 1966). The recent discovery of MSR in dolphins (Reiss and Marino 2001) thus fits the proposed connection between increased self-awareness, on the one hand, and cognitive empathy, on the other.

4d. Russian Doll

Stephanie Preston and de Waal (2002b) propose that at the core of the empathic capacity is the reactivation of the subject's stored representations of previously experienced states similar to those perceived in the object. This process relies on the subject's experience with these particular states as well as its closeness to the object. As a result, bonded individuals will respond more strongly to each other than socially distant individuals. This Perception-Action Model (PAM) fits Antonio Damasio's (1994) somatic marker hypothesis of emotions as well as recent evidence for a link at the cellular level between perception and action (e.g., "mirror neurons": di Pellegrino et al. 1992). The idea that perception and action share representations is anything but new: it goes as far back as the first treatise on *Einfühlung,* the German concept later translated into English as "empathy" (Wispé 1991). When T. Lipps (1903) introduced *Einfühlung*, which literally means "feeling into," he speculated about *innere Nachahmung* (inner mimicry) of another's feelings along the same lines as proposed by the PAM.

Empathy is often an insuppressible, unconscious process, as demonstrated by electromyographic studies of invisible muscle contractions in people's faces in response to pictures of human facial expressions that have been shown so briefly that subjects are unaware of them (e.g., Dimberg et al. 2000). Accounts of empathy as a higher cognitive process of simulation or perspective-taking, such as theory-of-mind, neglect these automatic "gut level" reactions, which are too rapid to depend on higher

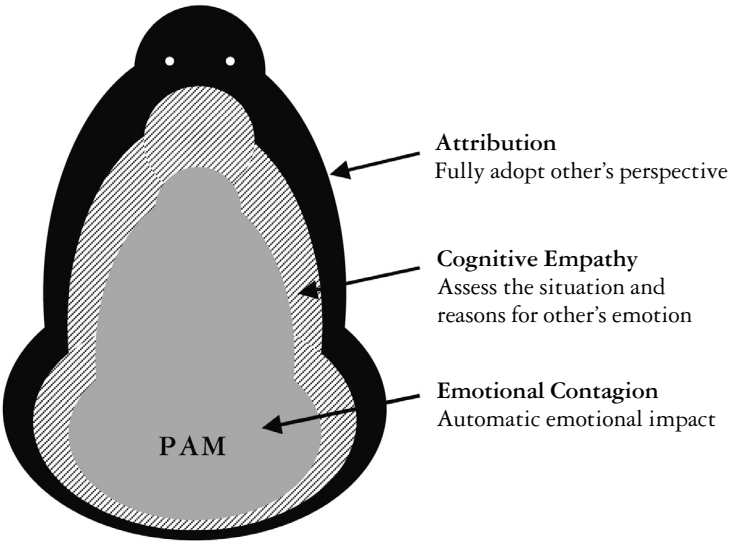

Figure 4. According to the Russian Doll Model, empathy covers all processes leading to related emotional states in subject and object; at its core are the perception-action mechanism (PAM) and emotional contagion: immediate, often unconscious state matching between individuals. Higher levels of empathy build on this hard-wired socio-affective basis, such as cognitive empathy (requiring an understanding of the reasons for the other's emotions) and mental state attribution (fully adopting the other's perspective). The Russian Doll Model proposes that these outer layers cannot exist without the inner ones. After de Waal (2003).

processes. This is not to say that more complex cognitive levels of empathy are irrelevant, yet they are built on top of this firm, hard-wired basis without which we would be at a loss about what moves others. This bottom-up view of empathy has led me to formulate a Russian doll model of empathy with a simple PAM-like mechanism at its core (figure 4).

Evolution never replaces anything: it works with existing structures and capacities, elaborating them and taking off from them. It has added to the PAM core a series of additions with increasingly complex components, such as emotional contagion, cognitive empathy, and attribution. Cognitive empathy implies appraisal of another's predicament or situation (cf. de Waal 1996). The subject not only responds to the signals emitted by the object but seeks to understand the reasons, looking for clues in the other's behavior and circumstances. Cognitive empathy

makes it possible to furnish targeted help that takes the needs of the other into account. These responses go well beyond emotional contagion, yet they would be hard to explain without an emotional motivational component.

Whereas monkeys (and most other social mammals) clearly seem to possess emotional contagion and some forms of targeted helping, the latter phenomenon is more strikingly developed in the great apes. That monkeys lack this capacity is evident from an example at Jigokudani Park in Japan, where first-time mother macaques are kept out of warm-water springs by park wardens because of their experience that these females tend to drown their infants accidentally. They fail to pay attention to them when submerging themselves in the ponds. This is some-

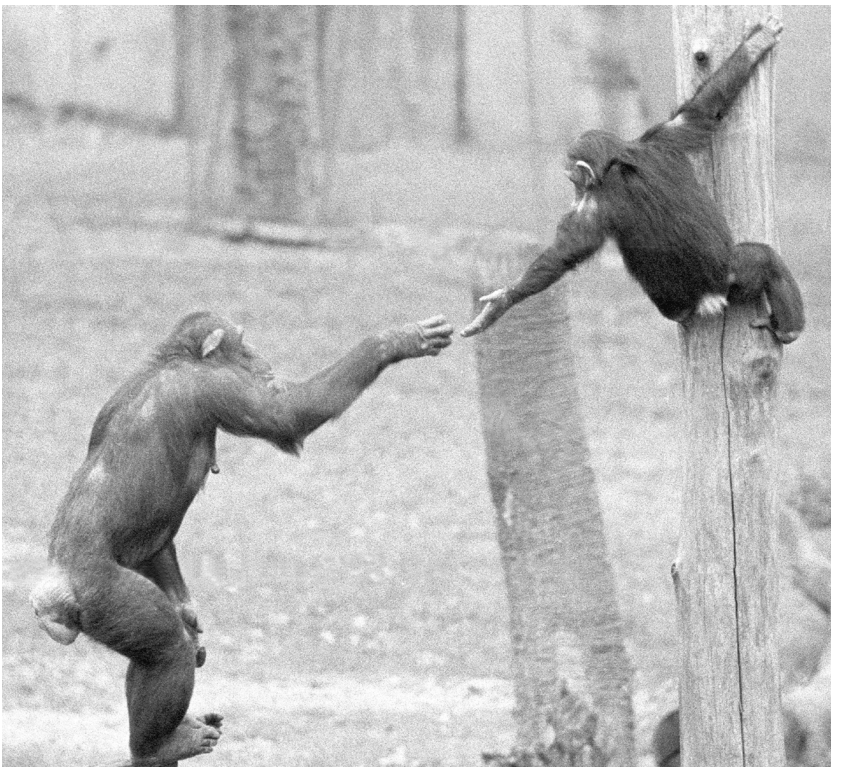

FIGURE 5. Cognitive empathy (i.e., empathy combined with appraisal of the other's situation) allows for aid tailored to the other's needs. In this case, a mother chimpanzee reaches out to help her son out of a tree after he has screamed and begged (see hand gesture). Targeted helping may require a distinction between self and other, an ability that may also underlie mirror self-recognition, as found in humans, apes, and dolphins. Photograph by the author.

thing they apparently learn over time, showing that they do not automatically take their offspring's perspective. De Waal (1996) speaks of "learned adjustment" to differentiate these acquired adaptive reactions from the immediate understanding shown by ape mothers, which tend to respond appropriately to specific needs of their offspring (figure 5).

In conclusion, empathy is not an all-or-nothing phenomenon: it covers a wide range of emotional linkage patterns, from the very simple and automatic to the very sophisticated. It seems logical first to try to understand the more basic forms, which are widespread indeed, before addressing the interesting variations that cognitive evolution has constructed on top of this foundation.

5. Reciprocity and Fairness

Chimpanzees and capuchin monkeys—the two species I work with most—are special, because they are among the very few primates that share food outside the mother-offspring context (Feistner and McGrew 1989). The capuchin is a small, easy primate to work with, as opposed to the chimpanzee, which is many times stronger than we are. Both species are interested in each other's food and will share food on occasion—sometimes even hand over a piece to another. Most sharing, however, is passive, where one individual will reach for food owned by another, who will let go. But even passive sharing is special compared to most animals, in which a similar situation would result in a fight or assertion by the dominant, without any sharing at all.

5*a. Chimpanzee Gratitude*

We studied sequences involving food sharing to see how a beneficial act by individual A toward B would affect B's behavior toward A. The prediction was that B would show beneficial behavior toward A in return. The problem with food sharing is, however, that after a group-wide feeding session as used in our experiments the motivation to share changes (the animals are more sated). Hence, food sharing cannot be the only variable measured. A second social service unaffected by food consumption was included. For this, grooming between individuals prior to food sharing was used. The frequency and duration of hundreds of spontaneous grooming bouts among our chimpanzees were measured in the morning. Within half an hour after the end of these observations, starting around noon, the apes were given two tightly bound bundles of leaves and branches. Nearly 7,000 interactions over food were carefully

recorded by observers and entered into a computer according to strict definitions described by de Waal (1989a). The resulting database on spontaneous services exceeds that for any other nonhuman primate.

It was found that adults were more likely to share food with individuals who had groomed them earlier. In other words, if A had groomed B in the morning, B was more likely than usual to share food with A later in the day. This result, however, could be explained in two ways. The first is the "good mood" hypothesis according to which individuals who have received grooming are in a benevolent mood, leading them to share indiscriminately with all individuals. The second explanation is the direct-exchange hypothesis, in which the individual who has been groomed responds by sharing food specifically with the groomer. The data indicated that the sharing increase was specific to the previous groomer. In other words, chimpanzees appeared to remember others who had just performed a service (grooming) and respond to those individuals by sharing more with them. Also, aggressive protests by food possessors to approaching individuals were directed more at those who had not groomed them than at previous grooming partners. This is compelling evidence for partner-specific reciprocal exchange (de Waal 1997b).

Of all existing examples of reciprocal altruism in nonhuman animals, the exchange of food for grooming in chimpanzees appears to be the most cognitively advanced. Our data strongly suggest a memory-based mechanism. A significant time delay existed between favors given and received (from half an hour to two hours); hence the favor was acted upon well after the previous interaction. Apart from memory of past events, we need to postulate that the memory of a received service, such as grooming, triggered a positive attitude toward the individual who offered the service, a psychological mechanism known as "gratitude" in humans. Gratitude in relation to the evolution of reciprocal exchange (cf. Trivers 1971) has been discussed at length by Kristin Bonnie and de Waal (2004) and is classified by Westermarck (1912 [1908]) as one of the "retributive kindly emotions" deemed essential for morality.

5*b. Monkey Fairness*

During the evolution of cooperation it may have become critical for actors to compare their own efforts and payoffs with those of others. Negative reactions may ensue in case of violated expectations. A recent theory proposes that aversion to inequity can explain human cooperation

within the bounds of the rational choice model (Fehr and Schmidt 1999). Similarly, cooperative nonhuman species seem guided by a set of expectations about the outcome of cooperation and access to resources. De Waal (1996: 95) proposed a *sense of social regularity,* defined as: "A set of expectations about the way in which oneself (or others) should be treated and how resources should be divided. Whenever reality deviates from these expectations to one's (or the other's) disadvantage, a negative reaction ensues, most commonly protest by subordinate individuals and punishment by dominant individuals."

The sense of how others should or should not behave is essentially egocentric, although the interests of individuals close to the actor, especially kin, may be taken into account (hence the parenthetical inclusion of others). Note that the expectations have not been specified: they are species-typical. To explore expectations held by capuchin monkeys we made use of their ability to judge and respond to value. We knew from previous studies that capuchins easily learn to assign value to tokens. Furthermore they can use these assigned values to complete a simple barter. This allowed a test to elucidate inequity aversion by measuring the reactions of subjects to a partner receiving a superior reward for the same tokens.

We paired each monkey with a group mate and watched their reactions when their partners got a better reward for doing the same bartering task. This consisted of an exchange in which the experimenter gave the subject a token that could immediately be handed back for a reward (figure 6). Each session consisted of twenty-five exchanges by each individual, and the subject always saw the partner's exchange immediately before its own. Food rewards varied from lower-value rewards (e.g., a cucumber piece), which they are usually happy to work for, to higher value rewards (e.g., a grape), which were preferred by all individuals tested. All subjects were subjected to (a) an Equity Test, in which subject and partner did the same work for the same low-value food, (b) an Inequity Test, in which the partner received a superior reward (grape) for the same effort, (c) an Effort Control Test, designed to elucidate the role of effort, in which the partner received the higher-value grape for free, and (d) a Food Control Test, designed to elucidate the effect of the presence of the reward on subject behavior, in which grapes were visible but not given to another capuchin.

Individuals who received lower-value rewards showed both passive negative reactions (e.g., refusing to exchange the token, ignoring the

FIGURE 6. A capuchin monkey in the test chamber returns a token to the experimenter with her right hand while steadying the human hand with her left hand. Her partner looks on. Drawing by Gwen Bragg and Frans de Waal after a video still.

reward) and active negative reactions (e.g., throwing out the token or the reward). Compared to tests in which both received identical rewards, the capuchins were far less willing to complete the exchange or accept the reward if their partner received a better deal (figure 7; Brosnan and de Waal 2003). Capuchins refused to participate even more frequently if their partner did not have to work (exchange) to get the better reward but was handed it for "free." Of course, there is always the possibility that subjects were just reacting to the presence of the higher-value food and that what the partner received (free or not) did not affect their reaction. However, in the Food Control Test, in which the higher-value reward was visible but not given to another monkey, the reaction to the presence of this high-valued food decreased significantly over the course of testing, which is a change in the opposite direction from that seen when the high-value reward went to an actual partner. Clearly our subjects discriminate between higher-value food being consumed by a conspecific and such food being merely visible, intensifying their rejections only to the former (Brosnan and de Waal 2004).

Capuchin monkeys thus seem to measure reward in relative terms, comparing their own rewards with those available and their own efforts with those of others. Although our data cannot elucidate the precise motivations underlying these responses, one possibility is that monkeys, like humans, are guided by social emotions. These emotions, known as "passions" by economists, guide human reactions to the efforts, gains, losses, and attitudes of others (Hirschleifer 1987; Frank 1988; Sanfey et

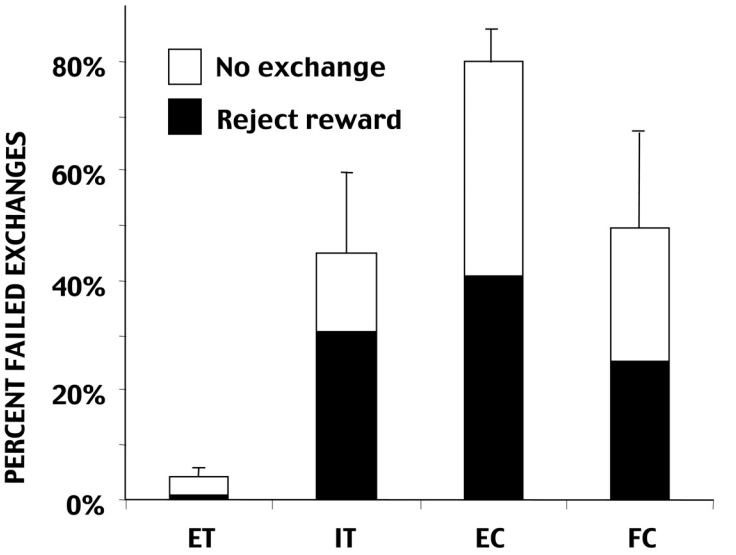

FIGURE 7. Mean percentage ± Standard Error of the Mean of failures to exchange for females across the four test types. Black bars represent the proportion of nonexchanges due to refusals to accept the reward; white bars represent nonexchanges due to refusals to return the token. ET = Equity Test, IT = Inequity Test, EC = Effort Control, FC = Food Control. The Y-axis shows the percentage of nonexchanges.

al. 2003). As opposed to primates marked by despotic hierarchies, tolerant species with well-developed food-sharing and cooperation, such as capuchin monkeys, may hold emotionally charged expectations about reward distribution and social exchange that lead them to dislike inequity.

Before we speak of "fairness" in this context it is good to point out a difference from human fairness. A full-blown sense of fairness would entail that the "rich" monkey shared with the "poor" one, as she would have felt she was getting excessive compensation for her efforts. Such behavior would have betrayed interest in a higher principle, one that Westermarck (1917 [1908]) called "disinterested," hence a truly moral notion. This is not the sort of reaction our monkeys showed, though; hence their sense of fairness, if we call it that, was more egocentric. They rather showed an expectation about how they themselves should be treated, not about how everybody around them should be treated. At the

same time, it cannot be denied that the full-blown sense of fairness must have started somewhere and that the self is the logical place to look for its origin. Once the egocentric form existed, it was expanded to include others.

6. Mencius and the Primacy of Affect

There is never much new under the sun. Westermarck's emphasis on the retributive emotions, whether friendly or vengeful, reminds one of the reply of Confucius to the question whether there is any single word that may serve as prescription for all of one's life. Confucius proposed "reciprocity" as such a word. Reciprocity is of course also at the heart of the Golden Rule, which remains unsurpassed as a summary of human morality. To know that some of the psychology behind this rule may exist in other species, along with the required empathy, bolsters the idea that morality, rather than being a recent invention, is part of human nature.

A follower of Confucius, Mencius, wrote extensively about human goodness during his life, from 372 to 289 BC. Mencius lost his father when he was three, and his mother made sure he received the best possible education. The mother is at least as well known as her son: she still serves as a maternal model to the Chinese for her absolute devotion to her son. Called the "second sage" because of his immense influence, second only to Confucius, Mencius had a revolutionary, subversive bent in that he stressed the obligation of rulers to provide for the common people. Recorded on bamboo clappers and handed down to his descendants and their students, his writings show that the debate about whether we are naturally moral or not is ancient indeed. In one exchange, Mencius (372–289 BC: 270–71) reacts against Kaou Tsze's views, which are astonishingly reminiscent of Huxley's gardener and garden metaphor:

> Man's nature is like the *ke* willow, and righteousness is like a cup or a bowl. The fashioning of benevolence and righteousness out of man's nature is like the making of cups and bowls from the *ke* willow.

Mencius replied:

> Can you, leaving untouched the nature of the willow, make with it cups and bowls? You must do violence and injury to the willow, before you can make cups and bowls with it. If you must do violence and injury to the willow, before you can make cups and bowls with it,

> *on your principles* you must in the same way do violence and injury to humanity in order to fashion from it benevolence and righteousness! Your words alas! would certainly lead all men on to reckon benevolence and righteousness to be calamities.

Mencius believed that humans tend toward the good as naturally as water flows downhill. This is also evident from the following remark, in which he seeks to exclude the possibility of the Freudian double-agenda on the grounds that the immediacy of the moral emotions, such as sympathy, leaves little room for cognitive contortions:

> When I say that all men have a mind which cannot bear to see the suffering of others, my meaning may be illustrated thus: even nowadays, if men suddenly see a child about to fall into a well, they will without exception experience a feeling of alarm and distress. They will feel so, not as a ground on which they may gain the favor of the child's parents, nor as a ground on which they may seek the praise of their neighbors and friends, nor from a dislike to the reputation of having been unmoved by such a thing. From this case we may perceive that the feeling of commiseration is essential to man. (Mencius 372–289 BC: 78)

This example from Mencius reminds us of the above epigraph from Westermarck ("Can we help sympathizing with our friends?") and the earlier quotation from Smith ("How selfish soever man may be supposed…"). The central idea underlying all three statements is that distress at the sight of another's pain is an impulse over which we exert no control: it grabs us instantaneously, like a reflex, without time to weigh the pros and cons. Remarkably, all alternative motives listed by Mencius occur in the modern literature, usually under the heading of reputation-building. The big difference is, of course, that Mencius rejected these explanations as too contrived, given the immediacy and force of the sympathetic impulse. Manipulation of public opinion is entirely possible at other times, he said, but not at the very instant that a child falls into a well.

I could not agree more. Evolution has produced species that follow genuinely cooperative impulses. I don't know if people are deep down good or evil, but I do know that to believe that each and every move is selfishly calculated while being hidden from ourselves and from others grossly overestimates human mental powers, let alone those of other animals. Apart from the already discussed animal examples of consolation

of distressed individuals and protection against aggression, there exists a rich literature on human empathy and sympathy that, generally, agrees with the assessment of Mencius that impulses in this regard come first and rationalizations later (e.g., Batson 1990; Wispé 1991).

7. Conclusion

In this lecture, I have drawn a stark contrast between two schools of thought on human goodness. One school sees people as essentially evil and selfish, and hence morality as a cultural overlay. This school of thought, personified by T. H. Huxley, is still very much with us even though I have noticed that no one (not even those explicitly endorsing this position) wants to be labeled a "veneer theorist." This is perhaps due to something about the wording itself, but also because once the assumptions of Veneer Theory are laid bare, it becomes obvious that the theory (a) lacks any sort of explanation of the transition from an amoral animal to a moral human being and (b) is at odds with empirical evidence bearing on moral judgment. If human morality truly operated entirely on the basis of calculations and rational decisions, without much emotional involvement, we would come close to being psychopaths, who indeed do not mean to be kind when they act kindly. Most of us hope to be slightly better than psychopaths; hence the widespread aversion to my black-and-white contrast between Veneer Theory and the other school, which seeks to ground morality in human nature.

This school sees morality arise naturally in our species and believes that there are sound evolutionary reasons for the capacities involved, even though it must be said that the theoretical framework to explain the transition from social to moral animal thus far consists of bits and pieces only. Its foundation is kin selection theory and reciprocal altruism, but other elements will need to be added. If one reads up on reputation building, fairness principles, empathy, and conflict resolution (disparate literatures that cannot be reviewed here in detail), there seems a promising movement toward a more integrated theory (for discussions, see Katz 2000).

According to this view of morality, the child is not going against its own nature by developing a caring, moral attitude any more than civil society is an out-of-control garden subdued by a sweating gardener. Moral attitudes have been with us from the beginning, and the gardener rather is, as John Dewey aptly put it, an organic grower. The successful gardener creates conditions and introduces plant species that may not be

normal for this particular plot of land "but fall within the wont and use of nature as a whole" (Dewey 1993 [1898]: 109–10). In other words we are not subduing the proverbial wolf within us or hypocritically fooling everyone around us when we act morally: we are taking decisions that flow from social instincts far older than our species, even though we add to these the perhaps uniquely human complexity of a disinterested concern for others and the society at large.

Following Hume (1978 [1739]), who saw reason as the slave of the passions, Jonathan Haidt (2001) has called for a thorough reevaluation of the role played by rationality in moral judgment, arguing that most human justification seems to occur *post hoc,* that is, after moral judgments have been reached on the basis of quick, automatic intuitions. A range of studies indicates the unconscious mirroring of others' emotional displays (Dimberg et al. 2000) and provides evidence for the PAM mechanism (i.e., activation of brain areas identical to those activated in the people with whom we empathize: see Preston and de Waal 2000b; Carr et al. 2003; Decety and Chaminade 2003; Wicker et al. 2003; Singer et al. 2004). Whereas Veneer Theory (which emphasizes human uniqueness in the domain of morality) would predict that moral considerations take place in evolutionarily recent additions to our brain, such as the neocortex, neuroimaging shows that moral judgments in fact involve a wide variety of brain areas, some rather ancient (Greene and Haidt 2002). In short, neuroscience is lending support to human morality as a relatively automated process closely tied to mammalian social instincts.

Additional evidence comes from child research. Developmental psychologists used to believe that the child learns its first moral distinctions through fear of punishment and a desire for praise. Similar to veneer theorists, they conceived morality as coming from the outside, imposed by adults upon a passive, naturally selfish child. Children were thought to adopt parental values to construct a superego, the moral agency of the self. Left to their own devices, children would never arrive at anything close to morality. We now know, however, that at an early age children understand the difference between moral principles ("do not steal") and cultural conventions ("no pajamas at school"). They apparently appreciate that the breaking of certain rules distresses and harms others, whereas the breaking of other rules merely violates expectations about what is appropriate. Their attitudes don't seem based purely on reward and punishment. Whereas many pediatric handbooks still depict young

children as self-centered monsters, it has become clear that by one year of age they spontaneously comfort people in distress (Zahn-Waxler et al. 1992) and that soon thereafter they begin to develop a moral perspective through interactions with other members of their species (Killen and Nucci 1995).

Instead of us doing "violence to the willow," as Mencius called it, to create the cups and bowls of an artificial morality, we rely on natural growth in which simple emotions, like those encountered in young children and social animals, develop into the more refined, other-including sentiments that we recognize as moral. My own argument obviously revolves around the continuity between human social instincts and those of our closest relatives, the monkeys and apes (de Waal 1996), but I feel that we are standing at the threshold of a much larger shift in theorizing that will end up positioning morality firmly within the emotional core of human nature.

Why did evolutionary biology stray from this path during the final quarter of the previous century? This is probably due to the conviction of some prominent figures, inspired by Huxley, that there is no way natural selection could have produced anything other than nasty organisms. No good could possibly have come from such a blind process. This belief, however, represents a monumental confusion between process and outcome. Natural selection is indeed a merciless process of elimination, yet it has the capacity to produce an incredible range of organisms, from the most asocial and competitive to the kindest and gentlest. If we assume that the building blocks of morality are among its many products, as Darwin did, then morality, instead of being a human-made veneer, should be looked at as an integral part of our history as group-living animals, hence an extension of our primate social instincts.

References

Alexander, R. A. 1987. *The Biology of Moral Systems.* New York: Aldine de Gruyter.

Arnhart, L. 1998. *Darwinian Natural Right: The Biological Ethics of Human Nature.* Albany, N.Y.: SUNY Press.

———. 1999. E. O. Wilson Has More in Common with Thomas Aquinas Than He Realizes. *Christianity Today International* 5 (6): 36.

Aureli, F., M. Cords, and C. P. van Schaik. 2002. Conflict Resolution Following Aggression in Gregarious Animals: A Predictive Framework. *Animal Behaviour* 64: 325–43.

Aureli, F., R. Cozzolino, C. Cordischi, and S. Scucchi. 1992. Kin-Oriented Redirection among Japanese Macaques: An Expression of a Revenge System? *Animal Behaviour* 44: 283–91.

Aureli, F., and F. B. M. de Waal. 2000. *Natural Conflict Resolution.* Berkeley: University of California Press.

Axelrod, R., and W. D. Hamilton. 1981. The Evolution of Cooperation. *Science* 211: 1390–96.

Badcock, C. R. 1986. *The Problem of Altruism: Freudian-Darwinian Solutions.* Oxford: Blackwell.

Bargh, J. A., and T. L. Chartrand. 1999. The Unbearable Automaticity of Being. *American Psychologist* 54: 462–79.

Baron-Cohen, S. 2000. Theory of Mind and Autism: A Fifteen Year Review. In *Understanding Other Minds*, ed. S. Baron-Cohen, H. Tager-Flusberg, and D. J. Cohen, pp. 3–20. Oxford: Oxford University Press.

Batson, C. D. 1990. How Social an Animal?: The Human Capacity for Caring. *American Psychologist* 45: 336–46.

Bischof-Köhler, D. 1988. Über den Zusammenhang von Empathie und der Fähigkeit sich im Spiegel zu erkennen. *Schweizerische Zeitschrift für Psychologie* 47: 147–59.

Boehm, C. 1999. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior.* Cambridge, Mass.: Harvard University Press.

Bonnie, K. E., and F. B. M. de Waal. 2004. Primate Social Reciprocity and the Origin of Gratitude. In *The Psychology of Gratitude*, ed. R. A. Emmons and M. E. McCullough, pp. 213–29. Oxford: Oxford University Press.

Brosnan, S. F., and F. B. M. de Waal. 2003. Monkeys Reject Unequal Pay. *Nature* 425: 297–99.

———. 2004. Reply to Commentators. *Nature* 428: 140.

Byrne, R. W., and A. Whiten. 1988. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans.* Oxford: Oxford University Press.

Caldwell, M. C., and D. K. Caldwell. 1966. Epimeletic (Care-Giving) Behavior in Cetacea. In *Whales, Dolphins, and Porpoises,* ed. K. S. Norris, pp. 755–89. Berkeley: University of California Press.

Carr, L., M. Iacoboni, M.-C. Dubeau, J. C. Mazziotta, and G. L. Lenzi. 2003. Neural Mechanisms of Empathy in Humans: A Relay from Neural Systems for Imitation to Limbic Areas. *Proceedings of the National Academy of Sciences* 100: 5497–5502.

Church, R. M. 1959. Emotional Reactions of Rats to the Pain of Others. *Journal of Comparative & Physiological Psychology* 52: 132–34.

Cohen, S., W. J. Doyle, D. P. Skoner, B. S. Rabin, and J. M. Gwaltney. 1997. Social Ties and Susceptibility to the Common Cold. *Journal of the American Medical Association* 277: 1940–44.

Damasio, A. 1994. *Descartes' Error: Emotion, Reason, and the Human Brain.* New York: Putnam.

Darwin, C. 1982 [1871]. *The Descent of Man, and Selection in Relation to Sex.* Princeton: Princeton University Press.

Dawkins, R. 1976. *The Selfish Gene.* Oxford: Oxford University Press.
——— 2003. *A Devil's Chaplain: Reflections on Hope, Lies, Science, and Love.* New York: Houghton Mifflin.
de Waal, F. B. M. 1989a. Food Sharing and Reciprocal Obligations among Chimpanzees. *Journal of Human Evolution* 18: 433–59.
———. 1989b. *Peacemaking among Primates.* Cambridge, Mass.: Harvard University Press.
———. 1996. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals.* Cambridge, Mass.: Harvard University Press.
———. 1997a. *Bonobo: The Forgotten Ape.* Berkeley: University of California Press.
———. 1997b. The Chimpanzee's Service Economy: Food for Grooming. *Evolution & Human Behavior* 18: 375–86.
———. 1998 [1982]. *Chimpanzee Politics: Power and Sex among Apes.* Baltimore, Md.: Johns Hopkins University Press.
———. 2000. Primates—A Natural Heritage of Conflict Resolution. *Science* 289: 586–90.
———. 2001. *The Ape and the Sushi Master: Cultural Reflections by a Primatologist.* New York: Basic Books.
———. 2003. On the Possibility of Animal Empathy. In *Feelings and Emotions: The Amsterdam Symposium,* ed. T. Manstead, N. Frijda, and A. Fischer, pp. 379–99. Cambridge: Cambridge University Press.
de Waal, F. B. M., and F. Aureli. 1996. Consolation, Reconciliation, and a Possible Cognitive Difference between Macaque and Chimpanzee. In *Reaching into Thought: The Minds of the Great Apes,* ed. A. E. Russon, K. A. Bard, and S. T. Parker, pp. 80–110. Cambridge: Cambridge University Press.
de Waal, F. B. M., and L. M. Luttrell. 1988. Mechanisms of Social Reciprocity in Three Primate Species: Symmetrical Relationship Characteristics or Cognition? *Ethology & Sociobiology* 9: 101–18.
de Waal, F. B. M., and A. van Roosmalen. 1979. Reconciliation and Consolation among Chimpanzees. *Behavioral Ecology & Sociobiology* 5: 55–66.
Decety, J., and T. Chaminade. 2003. Neural Correlates of Feeling Sympathy. *Neuropsychologia* 41: 127–38.
Desmond, A. 1994. *Huxley: From Devil's Disciple to Evolution's High Priest.* New York: Perseus.
Dewey, J. 1993 [1898]. Evolution and Ethics. Reprinted in *Evolutionary Ethics,* ed. M. H. Nitecki and D. V. Nitecki, pp. 95–110. Albany: State University of New York Press.
di Pellegrino, G., L. Fadiga, L. Fogassi, V. Gallese, and G. Rizzolatti. 1992. Understanding Motor Events: A Neurophysiological Study. *Experimental Brain Research* 91: 176–80.
Dimberg, U., M. Thunberg, and K. Elmehed. 2000. Unconscious Facial Reactions to Emotional Facial Expressions. *Psychological Science* 11: 86–89.
Dugatkin, L. A. 1997. *Cooperation among Animals: An Evolutionary Perspective.* New York: Oxford University Press.

Eisenberg, N. 2000. Empathy and Sympathy. In *Handbook of Emotion,* ed. M. Lewis and J. M. Haviland-Jones, pp. 677–91. 2nd ed. New York: Guilford Press.

Fehr, E., and K. M. Schmidt. 1999. A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics* 114: 817–68.

Feistner, A. T. C., and W. C. McGrew. 1989. Food-Sharing in Primates: A Critical Review. In *Perspectives in Primate Biology,* ed. P. K. Seth and S. Seth, vol. 3, pp. 21–36. New Delhi: Today & Tomorrow's Printers and Publishers.

Flack, J. C., and F. B. M. de Waal. 2000. "Any Animal Whatever": Darwinian Building Blocks of Morality in Monkeys and Apes. *Journal of Consciousness Studies* 7: 1–29.

Frank, R. H. 1988. *Passions within Reason: The Strategic Role of the Emotions.* New York: Norton.

Freud, S. 1913. *Totem and Taboo.* New York: Norton.

Gallup, G. G. 1982. Self-Awareness and the Emergence of Mind in Primates. *American Journal of Primatology* 2: 237–48.

Ghiselin, M. 1974. *The Economy of Nature and the Evolution of Sex.* Berkeley: University of California Press.

Gould, S. J. 1980. So Cleverly Kind an Animal. In *Ever Since Darwin,* pp. 260–67. Harmondsworth, UK: Penguin.

Greene, J., and J. Haidt. 2002. How (and Where) Does Moral Judgement Work? *Trends in Cognitive Sciences* 16: 517–23.

Haidt, J. 2001. The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review* 108: 814–34.

Harcourt, A. H., and F. B. M. de Waal. 1992 *Coalitions and Alliances in Humans and Other Animals.* Oxford: Oxford University Press.

Hatfield, E., J. T. Cacioppo, and R. L. Rapson. 1993. Emotional Contagion. *Current Directions in Psychological Science* 2: 96–99.

Hirschleifer, J. 1987. In *The Latest on the Best: Essays in Evolution and Optimality,* ed. J. Dupre, pp. 307–26. Cambridge, Mass.: MIT Press.

Hobbes, T. 1991 [1651]. *Leviathan.* Cambridge: Cambridge University Press.

Hoffman, M. L. 1981a. Is Altruism Part of Human Nature? *Journal of Personality and Social Psychology* 40: 121–37.

———. 1981b. Perspectives on the Difference between Understanding People and Understanding Things: The Role of Affect. In *Social Cognitive Development,* ed. J. H. Flavell and L. Ross, pp. 67–81. Cambridge: Cambridge University Press.

Hume, D. 1978 [1739]. *A Treatise of Human Nature.* Oxford: Oxford University Press.

Huxley, T. H. 1989 [1894]. *Evolution and Ethics.* Princeton: Princeton University Press.

Katz, L. D. 2000. *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives.* Exeter, UK: Imprint Academic.

Killen, M., and L. P. Nucci. 1995. Morality, Autonomy and Social Conflict. In *Morality in Everyday Life: Developmental Perspectives,* ed. M. Killen and D. Hart, pp. 52–86. Cambridge: Cambridge University Press.

Kropotkin, P. 1972 [1902]. *Mutual Aid: A Factor of Evolution.* New York: New York University Press.

Ladygina-Kohts, N. N. 2001 [1935]. *Infant Chimpanzee and Human Child: A Classic 1935 Comparative Study of Ape Emotions and Intelligence.* Ed. F. B. M. de Waal. New York: Oxford University Press.

Lipps, T. 1903. Einfühlung, innere Nachahmung und Organempfindung. *Archiv für die gesamte Psychologie* 1: 465–519.

Lorenz, K. Z. 1954. *Man Meets Dog.* London: Methuen.

———. 1966 [1963]. *On Aggression.* London: Methuen.

Masserman, J., M. S. Wechkin, and W. Terris. 1964. Altruistic Behavior in Rhesus Monkeys. *American Journal of Psychiatry* 121: 584–85.

Mayr, E. 1997. *This Is Biology: The Science of the Living World.* Cambridge, Mass.: Harvard University Press.

Mencius. 372–289 BC. *The Works of Mencius.* English translation by Gu Lu. Shanghai: Shangwu.

Midgley, M. 1979. Gene-Juggling. *Philosophy* 54: 439–58.

O'Connell, S. M. 1995. Empathy in Chimpanzees: Evidence for Theory of Mind? *Primates* 36: 397–410.

Preston, S. D., and F. B. M. de Waal. 2002a. The Communication of Emotions and the Possibility of Empathy in Animals. In *Altruistic Love: Science, Philosophy, and Religion in Dialogue,* ed. S. G. Post, L. G. Underwood, J. P. Schloss, and W. B. Hurlbut, pp. 284–308. Oxford: Oxford University Press.

——— 2002b. Empathy: Its Ultimate and Proximate Bases. *Behavioral & Brain Sciences* 25: 1–72.

Pusey, A. E., and C. Packer. 1987. Dispersal and Philopatry. In *Primate Societies,* ed. B. B. Smuts et al., pp. 250–66. Chicago: University of Chicago Press.

Rawls, J. 1972. *A Theory of Justice.* Oxford: Oxford University Press.

Reiss, D., and L. Marino. 2001. Mirror Self-Recognition in the Bottlenose Dolphin: A Case of Cognitive Convergence. *Proceedings of the National Academy of Science* 98: 5937–42.

Ridley, M. 1996. *The Origins of Virtue.* New York: Viking.

Roes, F. 1997. An Interview of Richard Dawkins. *Human Ethology Bulletin* 12 (1): 1–3.

Rothstein, S. I., and R. R. Pierotti. 1988. Distinctions among Reciprocal Altruism, Kin Selection, and Cooperation and a Model for the Initial Evolution of Beneficent Behavior. *Ethology & Sociobiology* 9: 189–209.

Sanfey, A. G., J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen. 2003. The Neural Basis of Economic Decision-making in the Ultimatum Game. *Science* 300: 1755–58.

Schleidt, W. M., and M. D. Shalter. 2003. Co-evolution of Humans and Canids, an Alternative View of Dog Domestication: *Homo homini lupus? Evolution and Cognition* 9: 57–72.

Silk, J. B., S. C. Alberts, and J. Altmann. 2003. Social Bonds of Female Baboons Enhance Infant Survival. *Science* 302: 1231–34.

Singer, T., B. Seymour, J. O'Doherty, K. Holger, R. J. Dolan, and C. D. Frith. 2004. Empathy for Pain Involves the Affective But Not Sensory Components of Pain. *Science* 303: 1157–62.

Smith, A. 1937 [1759]. *A Theory of Moral Sentiments.* New York: Modern Library.
Sober, E., and D. S. Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior.* Cambridge, Mass.: Harvard University Press.
Taylor, C. E., and M. T. McGuire. 1988. Reciprocal Altruism: Fifteen Years Later. *Ethology & Sociobiology* 9: 67–72.
Taylor, S. 2002. *The Tending Instinct.* New York: Times Books.
Todes, D. 1989. *Darwin without Malthus: The Struggle for Existence in Russian Evolutionary Thought.* New York: Oxford University Press.
Trivers, R. L. 1971. The Evolution of Reciprocal Altruism. *Quarterly Review of Biology* 46: 35–57.
van Schaik, C. P. 1983. Why Are Diurnal Primates Living in Groups? *Behaviour* 87: 120–44.
Watts, D. P., F. Colmenares, and K. Arnold. 2000. Redirection, Consolation, and Male Policing: How Targets of Aggression Interact with Bystanders. In *Natural Conflict Resolution,* ed. F. Aureli and F. B. M. de Waal, pp. 281–301. Berkeley: University of California Press.
Wechkin, S., J. H. Masserman, and W. Terris. 1964. Shock to a Conspecific as an Aversive Stimulus. *Psychonomic Science* 1: 47–48.
Westermarck, E. 1912 [1908]. *The Origin and Development of the Moral Ideas.* Vol. 1. 2nd ed. London: Macmillan.
———. 1917 [1908]. *The Origin and Development of the Moral Ideas.* Vol. 2. 2nd ed. London: Macmillan.
Wicker, B., C. Keysers, J. Plailly, J. P. Royet, V. Gallese, and G. Rizzolatti. 2003. Both of Us Disgusted in My Insula: The Common Neural Basis of Seeing and Feeling Disgust. *Neuron* 40: 655–64.
Williams, G. C. 1988. Reply to Comments on "Huxley's Evolution and Ethics in Sociobiological Perspective." *Zygon* 23: 437–38.
Wispé, L. 1991. *The Psychology of Sympathy.* New York: Plenum.
Wrangham, R. W. 1980. An Ecological Model of Female-Bonded Primate Groups. *Behaviour* 75: 262–300.
Wright, R. 1994. *The Moral Animal: The New Science of Evolutionary Psychology.* New York: Pantheon.
Yerkes, R. M. 1925. *Almost Human.* New York: Century.
Zahn-Waxler, C., B. Hollenbeck, and M. Radke-Yarrow. 1984. The Origins of Empathy and Altruism. In *Advances in Animal Welfare Science,* ed. M. W. Fox and L. D. Mickley, pp. 21–39. Washington, D.C.: Humane Society of the United States.
Zahn-Waxler, C., M. Radke-Yarrow, E. Wagner, and M. Chapman. 1992. Development of Concern for Others. *Developmental Psychology* 28: 126–36.
Zajonc, R. B. 1980. Feeling and Thinking: Preferences Need No Inferences. *American Psychologist* 35: 151–75.
———. 1984. On the Primacy of Affect. *American Psychologist* 39: 117–23.