THE UNIVERSITY CENTER

FOR HUMAN VALUES SERIES

STEPHEN MACEDO, EDITOR

Multiculturalism and "The Politics of Recognition"
by Charles Taylor

A Matter of Interpretation: Federal Courts and the Law
by Antonin Scalia

Freedom of Association edited by Amy Gutmann

Work and Welfare by Robert M. Solow

The Lives of Animals by J. M. Coetzee

Truth v. Justice: The Morality of Truth Commissions
edited by Robert I. Rotberg and Dennis Thompson

Goodness and Advice by Judith Jarvis Thomson

Human Rights as Politics and Idolatry
by Michael Ignatieff

Democracy, Culture and the Voice of Poetry
by Robert Pinsky

Primates and Philosophers: How Morality Evolved
by Frans de Waal

PRIMATES AND PHILOSOPHERS

How Morality Evolved

Frans de Waal

Robert Wright
Christine M. Korsgaard
Philip Kitcher
Peter Singer

EDITED AND INTRODUCED BY
Stephen Macedo and Josiah Ober

PRINCETON UNIVERSITY PRESS

PRINCETON AND OXFORD

Published by Princeton University Press, 41 William Street, Princeton,
New Jersey 08540

In the United Kingdom: Princeton University Press, 3 Market Place, Woodstock,
Oxfordshire OX20 1SY

Library of Congress Cataloging-in-Publication Data

Waal, F. B. M. de (Frans B. M.), 1948–
Primates and philosophers : how morality evolved / Frans de Waal ; edited and introduced by Stephen Macedo and Josiah Ober ; Christine M. Korsgaard . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-691-12447-6 (hardcover : alk. paper)
ISBN-10: 0-691-12447-7 (hardcover : alk. paper)
1. Ethics, Evolutionary. 2. Primates—Behavior. 3. Altruistic behavior in animals. I. Macedo, Stephen, 1957– II. Ober, Josiah. III. Korsgaard, Christine M. (Christine Marion) IV. Title.
BJ1311.W14 2006
171'.7—dc22
2006013905

British Library Cataloging-in-Publication Data is available

This book has been composed in Minion Family & Minion Condensed

Printed on acid-free paper.∞

pup.princeton.edu

Printed in the United States of America

5 7 9 10 8 6 4

Contents

PART I

MORALLY EVOLVED

PRIMATE SOCIAL INSTINCTS, HUMAN MORALITY, AND THE RISE AND FALL OF "VENEER THEORY"

Frans de Waal

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 wolf to man"—is an ancient Roman proverb popularized by Thomas Hobbes. Even though its basic tenet permeates large parts of law, economics, and political science, the proverb contains two major flaws. First, it fails to do justice to canids, which are among the most gregarious and cooperative animals on the planet (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* 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 strife became unbearable. According to Hobbes, social life

never came naturally to us. He saw it as a step we took reluctantly and "by covenant only, which is artificial" (Hobbes 1991 [1651]: 120). More recently, Rawls (1972) proposed a milder version of the same view, adding that humanity's move 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 evolution of our species. Hobbes and Rawls create the illusion of human society as a voluntary arrangement with self-imposed rules assented to by free and equal agents. Yet, there never was a point at which we became social: descended from highly social ancestors—a long line of 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. Any zoologist would classify our species as *obligatorily gregarious.*

Having companions offers immense advantages in locating food and avoiding predators (Wrangham 1980; van Schaik 1983). Inasmuch as group-oriented individuals leave more offspring than those less socially inclined (e.g., Silk et al. 2003), sociality has become ever more deeply ingrained in primate biology and psychology. If any decision to establish societies was made, therefore, credit should go to Mother Nature rather than to 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 *like* to live in. His 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, hence adopt it only for the sake of argument, it still distracts from the more pertinent argument that we ought to be pursuing, which is how we actually came to be what we are today. Which parts of human nature have led us down this 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, which is that we are social to the core.

A good illustration of the thoroughly 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 in the absence of others. We become hopelessly depressed 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 attachments as messy has made theoreticians turn to cognition as the preferred guide of human behavior. We celebrate rationality. 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).

Unfortunately, the emphasis on individual autonomy and rationality and a corresponding neglect of emotions and attachment are not restricted to the humanities and social sciences. Within evolutionary biology, too, some have embraced the notion 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 a direct outgrowth of the social instincts that we share with other animals. In the latter view, morality is neither unique to us nor a conscious decision taken at a specific point in time: it is the product of 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. Until recently, this was the dominant approach to morality within evolutionary biology as well as among science writers popularizing this field. I will use the term "Veneer Theory" to denote these ideas, tracing their origin to Thomas Henry Huxley (although they obviously go back much further in Western philosophy and religion, all the way to the concept

of original sin). 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 Westermarck, which agree with those of Darwin.

Given these contrasting opinions about continuity versus discontinuity with other animals, I then build upon an earlier treatise (de Waal 1996) in paying special attention to the behavior of nonhuman primates in order to explain why I think the building blocks of morality are evolutionarily ancient.

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 physical world are unalterable. He felt, however, that their impact on human existence could be softened and modified if people kept nature under control. Thus, Huxley compared humanity with a gardener who has a hard time keeping weeds out of his garden. He saw human ethics as a victory over an unruly and nasty 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 humanity, Huxley was in effect saying that what makes us human could not be handled by evolutionary theory. We can become moral only by opposing our own nature. This was an inexplicable retreat by someone who had gained a reputation as "Darwin's Bulldog" owing to his fierce advocacy of evolution. Second,

Huxley gave no hint whatsoever where humanity might have unearthed the will and strength to defeat the forces of 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? Would not this make us wolves in sheep's clothing: nice on the outside, nasty on the inside?

This was the only time Huxley 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 included morality in 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.

Huxley's curious dualism, which pits morality against nature and humanity 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 into 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 1962 [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 (Freud 1961 [1930]).

Humanity'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, Williams wrote 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, Dawkins waited until the very last sentence of *The Selfish Gene* to reassure us that, in fact, we are welcome to chuck all of those genes out 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. More recently, Dawkins (1996) has declared us "nicer than is good for our selfish genes," and explicitly endorsed 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; also Dawkins 2003).

Darwin must be turning in his grave, because the implied "Darwinian world" is miles removed from what he himself

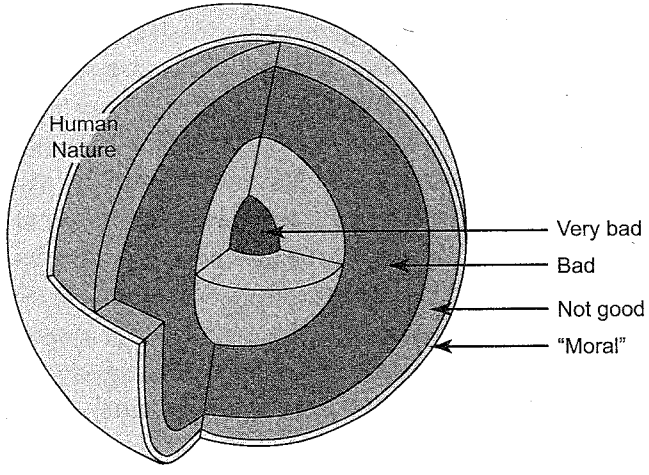

Figure 1 The popular view of morality among biologists during the past quarter of a century was summarized by Ghiselin (1974: 247): "Scratch an 'altruist,' and watch a 'hypocrite' bleed." Humans were considered thoroughly selfish and competitive, with morality being no more than an afterthought. Summarized as "Veneer Theory," this idea goes back to Darwin's contemporary, Thomas Henry Huxley. It is visualized here tongue-in-cheek as human nature bad to its core.

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 have depicted as all-powerful. Like the views of Hobbes, Huxley, and Freud, the thinking is thoroughly dualistic: we are part nature, part culture, rather than a well-integrated whole. Human morality is presented as a thin crust underneath of which boil antisocial, amoral, and egoistic passions. This view of morality as a veneer was best summarized by Ghiselin's famous quip: "Scratch an 'altruist,' and watch a 'hypocrite' bleed" (Ghiselin 1974: 247; figure 1).

Veneer Theory has since been popularized by countless science writers, such as Wright (1994), who went so far as to claim that virtue is absent from people's hearts and souls, and that our species is potentially but not naturally moral. One might ask: "But what about the people who occasionally experience in themselves and others a degree of sympathy, goodness, and generosity?" Echoing Ghiselin, Wright replies that the "moral animal" is essentially a hypocrite:

> [T]he 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 wishful 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)

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 effectively defended against its detractors. In the words of Mayr (1997: 250): "Huxley, who believed in final causes, rejected natural selection and did not represent genuine Darwinian thought in any way. . . . It is unfortunate, considering how confused Huxley was, that his essay [on ethics] is often referred to even today as if it were authoritative."

It should be pointed out, though, that in Huxley's time there was already fierce 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 an attack on Huxley, but written with great deference for Darwin.

Although Kropotkin never formulated his theory with the precision and evolutionary logic available to 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.

DARWIN ON 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 of morality 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 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 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-interests so long as one realizes that this by no means precludes the evolution of altruistic and sympathetic tendencies. Darwin fully 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 falling 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 probably first emerged with parental care, in which vulnerable individuals are fed and protected. In many animals they stretch beyond this domain, however, 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. This brings animal altruism much closer to that of humans than usually thought, and explains the call for the temporary removal of ethics from the hands of philosophers (Wilson 1975: 562).

Personally, I remain unconvinced that we need group selection to explain the origin of these tendencies—we seem to get quite far with the theories of kin selection and reciprocal altruism. Moreover, there is so much intergroup migration (hence gene flow) in nonhuman primates that the conditions for group selection do not seem fulfilled. In all of the primates, the younger generation of one sex or another (males in many monkeys, females in 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, which makes group selection unlikely.

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 seem relevant. Ants, too, share food, but likely based on quite different urges than those that make chimpanzees or people share (de Waal 1989a). This distinction was understood by Darwin, who looked beyond the actual behavior at the underlying emotions, 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 rather is whether they possess 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 profound misreading of Darwin. Since Darwin saw morality as an evolutionary product, he envisioned an eminently more livable

world than the one proposed by Huxley and his followers, who believe in a culturally imposed, artificial morality that receives no helping hand from human nature. Huxley's world is by far the colder, more terrifying place.

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 during his lifetime 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 at the time 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 camel driver 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 driver, "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 anticipating 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, but, as neuroscientists have found, 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 (1985 [1739]), 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 of the moral emotions were known—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 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 a 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 such 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.

Sections 4 and 5 discuss 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)—they are by no means sufficient

to produce morality as we know it, yet they are indispensable. No human moral society could be imagined without reciprocal exchange and an emotional interest in others. This offers a concrete starting point to investigate the continuity that Darwin envisioned. The debate about Veneer Theory is fundamental to this investigation since some evolutionary biologists have sharply deviated from the idea of continuity by presenting morality as a sham so convoluted that only one species—ours—is capable of it. This view has no basis in fact, and as such stands in the way of a full understanding of how we became moral (table 1). My intention here is to set the record straight by reviewing actual empirical data.

ANIMAL EMPATHY

Evolution rarely throws out anything. Structures are transformed, modified, co-opted for other functions, or "tweaked" in another direction—descent with modification, as Darwin called it. Thus, the frontal fins of fish became the front limbs of land animals, which over time turned into hoofs, paws, wings, hands, and flippers. Occasionally, a structure loses all function and becomes superfluous, but this is a gradual process, often ending in rudimentary traits rather than disappearance. We find tiny vestiges of leg bones under the skin of whales and remnants of a pelvis in snakes.

This is why to the biologist, a Russian doll is such a satisfying plaything, especially if it has a historical dimension. I own a doll that shows Russian President Vladimir Putin on the outside, within whom we discover, in this order, Yeltsin, Gorbachev, Brezhnev, Kruschev, Stalin, and Lenin. Finding a little Lenin and Stalin within Putin will hardly surprise most

TABLE 1
Comparison of Veneer Theory and the View of Morality as an Outgrowth of the Social Instincts

	Veneer Theory	*Evolution of Ethics*
Origin	Huxleyan	Darwinian
Advocates	Richard Dawkins, George Williams, Robert Wright, etc.	Edward Westermarck, Edward Wilson, Jonathan Haidt, etc.
Type	Dualistic—pits humans against animals, and culture against nature. Morality is seen as a choice.	Unitary—postulates continuity between human morality and animal social tendencies. Moral tendencies are seen as evolved.
Proposed transition	From amoral animal to moral human	From social to moral animal
Theory	A position in search of a theory. It offers no explanation of why humans are "nicer than is good for their selfish genes," nor how such a feat might have been accomplished.	Theories of kin selection, reciprocal altruism, and their derivatives (e.g., fairness, reputation building, conflict resolution)suggest how a transition from social to moral animal might have come about.
Empirical evidence	None	a) Psychology—human morality has an emotional and intuitive foundation. b) Neuroscience—moral dilemmas activate emotionally involved brain areas. c) Primate behavior—our relatives show many of the tendencies incorporated into human morality.

political analysts. The same is true for biological traits: the old always remains present in the new.

This is relevant to the debate about the origin of empathy, since the psychologist tends to look at the world through different eyes than the biologist. Psychologists sometimes put our most advanced traits on a pedestal, ignoring or even denying simpler antecedents. They thus believe in saltatory change, at least in relation to our own species. This leads to unlikely origin stories, postulating discontinuities with respect to language, which is said to result from a unique "module" in the human brain (e.g., Pinker 1994), or with respect to human cognition, which is viewed as having cultural origins (e.g., Tomasello 1999). True, human capacities reach dizzying heights, such as when I understand that you understand that I understand, et cetera. But we are not born with such "reiterated empathy," as phenomenologists call it. Both developmentally and evolutionarily, advanced forms of empathy are preceded by and grow out of more elementary ones. In fact, things may be exactly the other way around. Instead of language and culture appearing with a Big Bang in our species and then transforming the way we relate to each other, Greenspan and Shanker (2004) propose that it is from early emotional connections and "proto conversations" between mother and child (cf. Trevarthen 1993) that language and culture sprang. Instead of empathy being an endpoint, it may have been the starting point.

Biologists prefer bottom-up over top-down accounts, even though there is definitely room for the latter. Once higher order processes have come into existence, they modify processes at the base. The central nervous system is a good example of top-down processing, as in the control the prefrontal cortex exerts over memory. The prefrontal cortex is not the seat of

memory, but can "order" memory retrieval (Tomita et al. 1999). In the same way, culture and language shape expressions of empathy. The distinction between "being the origin of" and "shaping" is a fundamental one, though, and I will argue here that empathy is the original, pre-linguistic form of inter-individual linkage that only secondarily has come under the influence of language and culture.

Bottom-up accounts are the opposite of Big Bang theories. They assume continuity between past and present, child and adult, human and animal, even between humans and the most primitive mammals. We may assume that empathy first evolved in the context of parental care, which is obligatory in mammals (Eibl-Eibesfeldt 1974 [1971]; MacLean 1985). Signaling their state through smiling and crying, human infants urge their caregiver to pay attention and move into action (Bowlby 1958). The same applies to other primates. The survival value of these interactions is obvious. For example, a female chimpanzee lost a succession of infants despite intense positive interest because she was deaf and did not correct positional problems (such as sitting on the infant, or holding it upside-down) in response to its distress calls (de Waal 1998 [1982]).

For a human characteristic, such as empathy, that is so pervasive, develops so early in life (e.g., Hoffman 1975; Zahn-Waxler and Radke-Yarrow 1990), and shows such important neural and physiological correlates (e.g., Adolphs et al. 1994; Rimm-Kaufman & Kagan 1996; Decety and Chaminade 2003) as well as a genetic substrate (Plomin et al. 1993), it would be strange indeed if no evolutionary continuity existed with other mammals. The possibility of empathy and sympathy in other animals has been largely ignored, however. This is

partly due to an excessive fear of anthropomorphism, which has stifled research into animal emotions (Panksepp 1998; de Waal 1999, appendix A), and partly to the one-sided portrayal by biologists of the natural world as a place of combat rather than social connectedness.

What Is Empathy?

Social animals need to coordinate action and movement, collectively respond to danger, communicate about food and water, and assist those in need. Responsiveness to the behavioral states of conspecifics ranges from a flock of birds taking off all at once because one among them is startled by a predator to a mother ape who returns to a whimpering youngster to help it from one tree to the next by draping her body as a bridge between the two. The first is a reflex-like transmission of fear that may not involve any understanding of what triggered the initial reaction, but that is undoubtedly adaptive. The bird that fails to take off at the same instant as the rest of the flock may be lunch. The selection pressure on paying attention to others must have been enormous. The mother-ape example is more discriminating, involving anxiety at hearing one's offspring whimper, assessment of the reason for its distress, and an attempt to ameliorate the situation.

There exists ample evidence of one primate coming to another's aid in a fight, putting an arm around a previous victim of attack, or other emotional responses to the distress of others (to be reviewed below). In fact, almost all communication among nonhuman primates is thought to be emotionally mediated. We are familiar with the prominent role

of emotions in human facial expressions (Ekman 1982), but when it comes to monkeys and apes—which have a homologous array of expressions (van Hooff 1967)—emotions seem equally important.

When the emotional state of one individual induces a matching or closely related state in another, we speak of "emotional contagion" (Hatfield et al. 1993). Even if such contagion is undoubtedly a basic phenomenon, there is more to it than simply one individual being affected by the state of another: the two individuals often engage in direct interaction. Thus, a rejected youngster may throw a screaming tantrum at its mother's feet, or a preferred associate may approach a food possessor to beg by means of sympathy-inducing facial expressions, vocalizations, and hand gestures. In other words, emotional and motivational states often manifest themselves in behavior specifically directed at a partner. The emotional effect on the other is not a by-product, therefore, but actively sought.

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 have arisen without—emotional contagion, but it goes beyond it in that it places filters between the other's and one's own state. In humans, it is around the age of two that we begin to add these cognitive layers (Eisenberg and Strayer 1987).

Two mechanisms related to empathy are *sympathy* and *personal distress*, which in their social consequences are each other's opposites. Sympathy is defined as "an affective response that consists of feelings of sorrow or concern for a distressed or needy other (rather than the same emotion as

the other person). Sympathy is believed to involve an other-oriented, altruistic motivation" (Eisenberg 2000: 677). Personal distress, on the other hand, makes the affected party selfishly seek to alleviate its *own* distress, which is similar to what it has perceived in the object. Personal distress is therefore not concerned with the situation of the empathy-inducing other (Batson 1990). A striking primate example is given by de Waal (1996: 46): the screams of a severely punished or rejected infant rhesus monkey will often cause other infants to approach, embrace, mount, or even pile on top of the victim. Thus, the distress of one infant seems to spread to its peers, which then seek contact to soothe their own arousal. Inasmuch as personal distress lacks cognitive evaluation and behavioral complementarity, it does not reach beyond the level of emotional contagion.

That most modern textbooks on animal cognition (e.g., Shettleworth 1998) fail to index empathy or sympathy does not mean that these capacities are not an essential part of animal lives; it only means that they are being overlooked by a science traditionally focused on individual rather than inter-individual capacities. Tool use and numerical competence, for instance, are seen as hallmarks of intelligence, whereas appropriately dealing with others is not. It is obvious, however, that survival often depends on how animals fare within their group, both in a cooperative sense (e.g., concerted action, information transfer) and in a competitive sense (e.g., dominance strategies, deception). It is in the *social* domain, therefore, that one expects the highest cognitive achievements. Selection must have favored mechanisms to evaluate the emotional states of others and quickly respond to them. Empathy is precisely such a mechanism.

In human behavior, there exists a tight relation between

empathy and sympathy, and their expression in psychological altruism (e.g., Hornblow 1980; Hoffman 1982; Batson et al. 1987; Eisenberg and Strayer 1987; Wispé 1991). It is reasonable to assume that the altruistic and caring responses of other animals, especially mammals, rest on similar mechanisms. When 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. Since expressions of sympathy emerge at an early age in virtually every member of our species, they are as natural as the first step. An unplanned sidebar to this 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).

Rooted in attachment and what Harlow termed the "affectional system" (Harlow and Harlow 1965), responses to the emotions of others are commonplace in social animals. Thus, behavioral and physiological data suggest emotional contagion in a variety of species (reviewed in Preston and de Waal 2002b, and de Waal 2003). An interesting literature that appeared in the 1950s and '60s by experimental psychologists placed the words "empathy" and "sympathy" between quotation marks. In those days, talk of animal emotions was taboo. In a paper provocatively entitled "Emotional Reactions of Rats to the Pain of Others," 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 witness them.

Monkeys show a stronger inhibition than rats. The most compelling evidence for the strength of empathy in monkeys came from Wechkin et al. (1964) and 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 these macaques, as supported by the finding that the inhibition to hurt another was more pronounced between familiar than unfamiliar individuals (Masserman et al. 1964).

Although these early studies suggest that, by behaving in certain ways, animals try to alleviate or prevent distress in others, it remains unclear if spontaneous responses to distressed conspecifics are explained by (a) aversion to distress signals of others, (b) personal distress generated through emotional contagion, or (c) true helping motivations. Work on nonhuman primates has furnished further information. Some of this evidence is qualitative, but quantitative data on empathic reactions exists as well.

Anecdotes of "Changing Places in Fancy"

Striking depictions of primate empathy and altruism can be found in Yerkes (1925), Ladygina-Kohts (2002 [1935]), Goodall (1990), and de Waal (1998 [1982], 1996, 1997a). Primate empathy is such a rich area that O'Connell (1995)

was able to conduct a content analysis of thousands of qualitative reports. She concluded that responses to the distress of another seem considerably more complex in apes than monkeys. To give just one example of the strength of the ape's empathic response, Ladygina-Kohts wrote about her young chimpanzee, Joni, that the best way to get him off the roof of her house (much better than any reward or threat of punishment) was by arousing 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. (Ladygina-Kohts, 2002 [1935]: 121)

De Waal (1996, 1997a) has suggested that apart from emotional connectedness, apes have an appreciation of the other's situation and a degree of perspective-taking (appendix B). So, the main difference between monkeys and apes is not in empathy *per se*, but in the cognitive overlays, which allow apes to adopt the other's viewpoint. One 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, p. 156)

What Kuni did would obviously have been inappropriate towards 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 offering us an anthropoid version of the empathic capacity so enduringly described by Adam Smith (1937 [1759]: 10) as "changing places in fancy with the sufferer." Perhaps the most striking example of this capacity is a chimpanzee who, as in the original Theory of Mind (ToM) experiments of Premack and Woodruff (1978), seemed to understand the intentions of another and provided specific assistance:

> During one winter at the Arnhem Zoo, after cleaning the hall and before releasing the chimps, the keepers hosed out all rubber tires in the enclosure and hung them one by one on a horizontal log extending from the climbing frame. One day, Krom was interested in a tire in which water had stayed behind. Unfortunately, this particular tire was at the end of the row, with six or more heavy tires hanging in front of it. Krom pulled and pulled at the one she wanted but couldn't remove it from the log. She pushed the tire backward, but there it hit the climbing frame and couldn't be removed

either. Krom worked in vain on this problem for over ten minutes, ignored by everyone, except Jakie, a seven-year-old Krom had taken care of as a juvenile.

Immediately after Krom gave up and walked away, Jakie approached the scene. Without hesitation he pushed the tires one by one off the log, beginning with the front one, followed by the second in the row, and so on, as any sensible chimp would. When he reached the last tire, he carefully removed it so that no water was lost, carrying it straight to his aunt, placing it upright in front of her. Krom accepted his present without any special acknowledgment, and was already scooping up water with her hand when Jakie left. (Adapted from de Waal 1996)

That Jakie assisted his aunt is not so unusual. What is special is that he correctly guessed what Krom was after. He grasped his auntie's goals. Such so-called "targeted helping" is typical of apes, but rare or absent in most other animals. It is defined as altruistic behavior tailored to the specific needs of the other even in novel situations, such as the highly publicized case of Binti Jua, a female gorilla who rescued a human child at the Brookfield Zoo in Chicago (de Waal, 1996, 1999). A recent experiment demonstrated targeted helping in young chimpanzees (Warneken and Tomasello 2006).

It is important to stress the incredible strength of the ape's helping response, which makes these animals take great risks on behalf of others. Whereas in a recent debate about the origins of morality, Kagan (2000) considered it obvious that a chimpanzee would never jump into a cold lake to save another, it may help to quote Goodall (1990: 213) on this issue:

In some zoos, chimpanzees are kept on man-made islands, surrounded by water-filed moats. . . . Chimpanzees cannot swim and, unless they are rescued, will drown if they fall into deep water. Despite this, individuals have sometimes made heroic efforts to save companions from drowning—and were sometimes successful. One adult male lost his life as he tried to rescue a small infant whose incompetent mother had allowed it to fall into the water.

The only other animals with a similar array of helping responses are dolphins and elephants. This evidence, too, is largely descriptive (dolphins: Caldwell and Caldwell 1966; Connor and Norris 1982; elephants: Moss 1988; Payne 1998), yet here again it is hard to accept as coincidental that scientists who have watched these animals for any length of time have numerous such stories, whereas scientists who have watched other animals have few, if any.

Consolation Behavior

This difference between monkey and ape empathy has been confirmed by systematic studies of a behavior known as "consolation," first documented by de Waal and van Roosmalen (1979). Consolation is defined as reassurance 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 his or her shoulders (figure 2). Consolation is not to be confused with reconciliation between former opponents, which seems mostly motivated by self-interest, such as the imperative to restore a disturbed social relationship (de Waal 2000). The advantage of consolation for the actor remains wholly

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.

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 by de Waal and 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

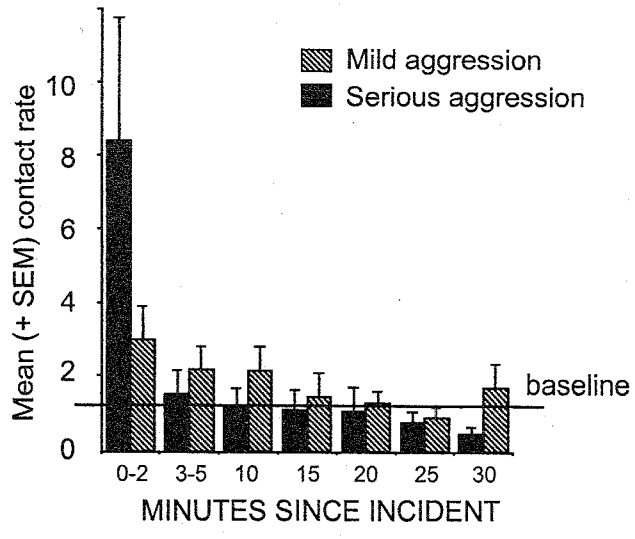

Figure 3 The rate at which third parties contact victims of aggression in chimpanzees, comparing recipients of serious and mild aggression. Especially in the first few minutes after the incident, recipients of serious aggression receive more contacts than baseline. After de Waal and Aureli (1996).

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 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 data

collection method, have shown reconciliation in species after species. Why, then, would consolation be restricted to apes?

Possibly, one cannot achieve cognitive empathy without a high degree of self-awareness. 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. In other words, in order to understand that the source of vicarious arousal is not oneself but the other and to understand the causes of the other's state, one needs a clear distinction between self and other. Based on these assumptions, Gallup (1982) was the first to speculate about a 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 (i.e., humans and apes) but not 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 Binti Jua's rescue of a boy. These responses require an understanding of the specific predicament of the individual needing help. Given the evidence for targeted helping by dolphins (see above), the recent discovery of MSR in these mammals (Reiss and Marino 2001) supports the proposed connection between increased self-awareness, on the one hand, and cognitive empathy, on the other.

Russian Doll Model

The literature includes accounts of empathy as a cognitive affair, even to the point that apes, let alone other animals, probably lack it (Povinelli 1998; Hauser 2000). This view equates empathy with mental state attribution and ToM. The opposite position has recently been defended in relation to autistic children, however. Contra earlier assumptions that autism reflects a ToM deficit (Baron-Cohen 2000), autism is noticeable well before the age of 4 years at which ToM typically emerges. Williams et al. (2001) argue that the main deficit of autism concerns the socio-affective level, which in turn negatively impacts sophisticated downstream forms of interpersonal perception, such as ToM. Thus, ToM is seen as a derived trait, and the authors urge more attention to its antecedents (a position now also embraced by Baron-Cohen 2003, 2004).

Preston and de Waal (2002a) propose that at the core of the empathic capacity is a relatively simple mechanism that provides an observer (the "subject") with access to the emotional state of another (the "object") through the subject's own neural and bodily representations. When the subject attends to the object's state, the subject's neural representations of similar states are automatically activated. The closer and more similar subject and object are, the easier it will be for the subject's perception to activate motor and autonomic responses that match the object's (e.g., changes in heart rate, skin conductance, facial expression, body posture). This activation allows the subject to get "under the skin" of the object, sharing its feelings and needs, which embodiment in turn fosters sympathy, compassion, and helping. Preston

and de Waal's (2002a) Perception-Action Mechanism (PAM) fits 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 Pelligrino 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 translated into English as "empathy" (Wispé 1991). When Lipps (1903) spoke of *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. Accordingly, empathy is a routine involuntary process, as demonstrated by electromyographic studies of invisible muscle contractions in people's faces in response to pictures of human facial expressions. These reactions are fully automated and occur even when people are unaware of what they saw (Dimberg et al. 2000). Accounts of empathy as a higher cognitive process neglect these gut-level reactions, which are far too rapid to be under conscious control.

Perception-action mechanisms are well known for motor perception (Prinz and Hommel 2002), causing researchers to assume similar processes to underlie emotion perception (Gallese 2001; Wolpert et al. 2001). Data suggest that both observing and experiencing emotions involves shared physiological substrates: *seeing* another's disgust or pain is very much like *being* disgusted or in pain (Adolphs et al. 1997, 2000; Wicker et al. 2003). Also, affective communication creates similar physiological states in subject and object (Dimberg 1982, 1990; Levenson and Reuf 1992). In short, human physiological and neural activity does not take place on an island, but is intimately connected with and affected by

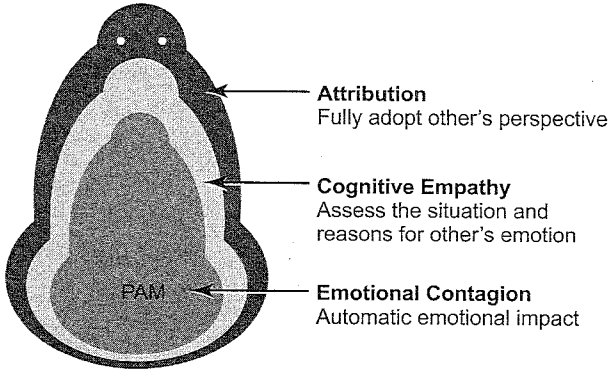

Figure 4 According to the Russian Doll Model, empathy covers all processes leading to related emotional states in subject and object. At its core is a simple, automatic Perception-Action Mechanism (PAM), which results in immediate, often unconscious state matching between individuals. Higher levels of empathy that build on this hardwired basis include cognitive empathy (i.e., understanding the reasons for the other's emotions) and mental state attribution (i.e., fully adopting the other's perspective). The Russian Doll Model proposes that outer layers require inner ones. After de Waal (2003).

fellow human beings. Recent investigations of the neural basis of empathy lend strong support to the PAM (Carr et al. 2003; Singer et al. 2004; de Gelder et al. 2004).

How simple forms of empathy relate to more complex ones has been depicted as a Russian doll by de Waal (2003). Accordingly, empathy covers all forms of one individual's emotional state affecting another's, with basic mechanisms at its core and more advanced mechanisms and cognitive abilities as its outer layers (figure 4). Autism may be reflected in deficient outer layers of the Russian doll, but such deficiencies invariably go back to deficient inner layers.

This is not to say that higher cognitive levels of empathy

are irrelevant, but they are built on top of this firm, hardwired basis without which we would be at a loss about what moves others. Surely, not all empathy is reducible to emotional contagion, but it never gets around it. At the core of the Russian doll, we find a PAM-induced emotional state that corresponds with the object's state. In a second layer, 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 for these signals, looking for clues in the other's behavior and situation. Cognitive empathy makes it possible to furnish targeted help that takes the specific needs of the other into account (figure 5). These responses go well beyond emotional contagion, yet they would be hard to explain without the motivation provided by the emotional component. Without it, we would be as disconnected as Mr. Spock in *Star Trek*, constantly wondering why others feel what they say they feel.

Whereas monkeys (and many other social mammals) clearly seem to possess emotional contagion and a limited degree of targeted helping, the latter phenomenon is not nearly as robust as in the great apes. For example, at Jigokudani Monkey Park, in Japan, first-time mother macaques are kept out of the hot springs by park wardens because of the experience that these females will accidentally drown their infants. They fail to pay attention to them when submerging themselves in the ponds. This is something monkey mothers apparently have to learn with time, showing that they do not automatically take their offspring's perspective. De Waal (1996) ascribed their behavioral change to "learned adjustment," setting it apart from cognitive empathy, which is more typical of apes and humans. Ape mothers respond immediately and appropriately to the specific needs of their

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 also thought to underlie mirror self-recognition, as found in humans, apes, and dolphins. Photograph by the author.

offspring. They are, for example, very careful to keep them away from water, rushing over to pull them away as soon as they get too close.

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 highly sophisticated. It seems logical to first try to understand the basic

forms of empathy, which are widespread indeed, before addressing the variations that cognitive evolution has constructed on top of this foundation.

RECIPROCITY AND FAIRNESS

Chimpanzees and capuchin monkeys—the two species I work with most—are special, as they are among the very few primates that share food outside the mother-offspring context (Feistner and McGrew 1989). The capuchin is a small primate, easy to work with, as opposed to the chimpanzee, which is many times stronger than we are. Members of 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 when compared to most animals, for which a similar situation would result in a fight or assertion by the dominant, without any sharing at all.

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 this service, a psychological mechanism known in humans as "gratitude." Gratitude within the context of reciprocal exchange was predicted by Trivers (1971), and has been discussed by Bonnie and de Waal (2004). It is classified by Westermarck (1912 [1908]) as one of the "retributive kindly emotions" deemed essential for human morality.

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 tend to be species-typical. For example, a rhesus monkey expects no share of a dominant's food, as it lives in a despotically hierarchical society, but a chimpanzee definitely does, hence the begging, whining, and temper tantrums if no share is forthcoming. I consider expectations the most important unstudied topic in animal behavior, which is all the more lamentable as it is the one issue that will bring animal behavior closest to the "ought" of behavior that we recognize so clearly in the moral domain.

To explore the 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,

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.

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 (ET), in which subject and partner did the same work for the same low-value food, (b) an Inequity Test (IT), in which the partner received a superior reward (grape) for the same effort, (c) an Effort Control Test (EC), designed to elucidate the role of effort, in

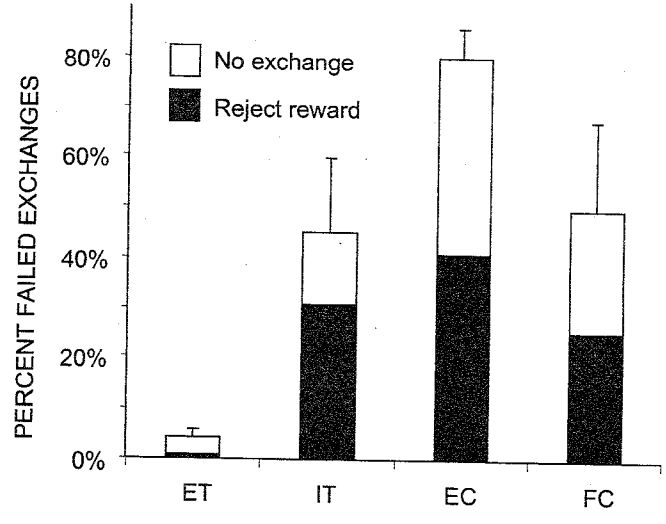

Figure 7 Mean percentage ± standard error of the mean of failures to exchange for females across the four test types. Black bars represent the proportion of nonexchanges due to refusals to accept the reward; white bars represent nonexchanges due to refusals to return the token. ET = Equity Test, IT =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

which the partner received the higher value grape for free, and (d) a Food Control Test (FC), 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 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; Bros-

nan 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-value 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 2003).

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. In humans, these emotions, known as "passions" by economists, guide an individual's reactions to the efforts, gains, losses, and attitudes of others (Hirschleifer 1987; Frank 1988; Sanfey et al. 2003). As opposed to primates marked by despotic hierarchies (such as rhesus monkeys), 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 between this and human fairness,

though. A full-blown sense of fairness would entail that the "rich" monkey share with the "poor" one, as she should feel she is getting excessive compensation. Such behavior would betray interest in a higher principle of fairness, one that Westermarck (1917 [1908]) called "disinterested," hence a truly moral notion. This is not the sort of reaction our monkeys showed, though: their sense of fairness, if we call it that, was rather egocentric. They 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 someplace and that the self is the logical place to look for its origin. Once the egocentric form exists, it can be expanded to include others.

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 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: to the Chinese, she still serves as a maternal model for her absolute devotion. 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 (n.d. [372–289 BC]: 270–71) reacts against Kaou Tsze's views, which are 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 between presented and felt

motives on the grounds that the immediacy of the moral emotions, such as sympathy, leaves no 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 n.d. [372–289 BC]: 78)

This example from Mencius reminds us of Westermarck's epigraph ("Can we help sympathizing with our friends?") and the 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 little or no control: it grabs us instantaneously, like a reflex, without time to weigh the pros and cons. All three statements hint at an involuntary process such as PAM. Remarkably, the possible alternative motives brought up by Mencius also feature 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 to believe that each and every move is selfishly calculated, while being hidden from others (and often from ourselves), seems to grossly overestimate human intellectual 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).

COMMUNITY CONCERN

In this essay, 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 mere cultural overlay. This school, 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) likes to be called a "veneer theorist." This may be due to wording, or because once the assumptions behind Veneer Theory are laid bare, it becomes obvious that—unless one is willing to go the purely rationalist route of modern Hobbesians, such as Gauthier (1986)—the theory lacks any sort of explanation of how we moved from being amoral animals to moral beings. The theory is at odds with the evidence for emotional processing as driving force behind moral judgment. If human morality could truly be reduced to calculations and reasoning, 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 that, hence the possible aversion to my black-and-white contrast between Veneer Theory and the alternative 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. Nevertheless, the theoretical framework to explain the transition from social animal to moral human consists only of bits and pieces. Its foundations are the theories of kin selection and reciprocal altruism, but it is obvious that other elements will need to be added. If one reads up on reputation building, fairness principles, empathy, and conflict resolution (in disparate literatures that cannot be reviewed here), there seems a promising movement toward a more integrated theory of how morality may have come about (see Katz 2000).

It should further be noted that the evolutionary pressures responsible for our moral tendencies may not all have been nice and positive. After all, morality is very much an in-group phenomenon. Universally, humans treat outsiders far worse than members of their own community: in fact, moral rules hardly seem to apply to the outside. True, in modern times there is a movement to expand the circle of morality, and to include even enemy combatants—e.g., the Geneva Convention, adopted in 1949—but we all know how fragile an effort this is. Morality likely evolved as a within-group phenomenon in conjunction with other typical within-group capacities, such as conflict resolution, cooperation, and sharing.

The first loyalty of every individual is not to the group, however, but to itself and its kin. With increasing social integration and reliance on cooperation, shared interests

must have risen to the surface so that the community as a whole became an issue. The biggest step in the evolution of human morality was the move from interpersonal relations to a focus on the greater good. In apes, we can see the beginnings of this when they smooth relations between others. Females may bring males together after a fight between them, thus brokering a reconciliation, and high-ranking males often stop fights among others in an evenhanded manner, thus promoting peace in the group. I see such behavior as a reflection of *community concern* (de Waal 1996), which in turn reflects the stake each group member has in a cooperative atmosphere. Most individuals have much to lose if the community were to fall apart, hence the interest in its integrity and harmony. Discussing similar issues, Boehm (1999) added the role of social pressure, at least in humans: the entire community works at rewarding group-promoting behavior and punishing group-undermining behavior.

Obviously, the most potent force to bring out a sense of community is enmity toward outsiders. It forces unity among elements that are normally at odds. This may not be visible at the zoo, but it is definitely a factor for chimpanzees in the wild, which show lethal intercommunity violence (Wrangham and Peterson 1996). In our own species, nothing is more obvious than that we band together against adversaries. In the course of human evolution, out-group hostility enhanced in-group solidarity to the point that morality emerged. Instead of merely ameliorating relations around us, as apes do, we have explicit teachings about the value of the community and the precedence it takes, or ought to take, over individual interests. Humans go much further in all of this than the apes (Alexander 1987), which is why we have moral systems and apes do not.

And so, the profound irony is that our noblest achievement—morality—has evolutionary ties to our basest behavior—warfare. The sense of community required by the former was provided by the latter. When we passed the tipping point between conflicting individual interests and shared interests, we ratcheted up the social pressure to make sure everyone contributed to the common good.

If we accept this view of an evolved morality, of morality as a logical outgrowth of cooperative tendencies, we are not going against our own nature by developing a caring, moral attitude, any more than civil society is an out-of-control garden subdued by a sweating gardener, as Huxley (1989 [1894]) thought. Moral attitudes have been with us from the start, and the gardener rather is, as 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 hypocritically fooling everyone when we act morally: we are making decisions that flow from social instincts older than our species, even though we add to these the uniquely human complexity of a disinterested concern for others and for society as a whole.

Following Hume (1985 [1739]), who saw reason as the slave of the passions, 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, automated intuitions. Whereas Veneer Theory, with its emphasis on human uniqueness, would predict that moral problem solving is assigned to evolutionarily recent additions to our brain, such as the prefrontal

cortex, neuroimaging shows that moral judgment in fact involves a wide variety of brain areas, some extremely ancient (Greene and Haidt 2002). In short, neuroscience seems to be lending support to human morality as evolutionarily anchored in mammalian sociality.

We celebrate rationality, but when push comes to shove we assign it little weight (Macintyre 1999). This is especially true in the moral domain. Imagine that an extraterrestrial consultant instructs us to kill people as soon as they come down with influenza. In doing so, we are told, we would kill far fewer people than would die if the epidemic were allowed to run its course. By nipping the flu in the bud, we would save lives. Logical as this may sound, I doubt that many of us would opt for this plan. This is because human morality is firmly anchored in the social emotions, with empathy at its core. Emotions are our compass. We have strong inhibitions against killing members of our own community, and our moral decisions reflect these feelings. For the same reasons, people object to moral solutions that involve hands-on harm to another (Greene and Haidt 2002). This may be because hands-on violence has been subject to natural selection, whereas utilitarian deliberations have not.

Additional support for an intuitionist approach to morality 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 know now, 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 others 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 our 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 underlying morality. My own argument here obviously revolves around the continuity between human social instincts and those of our closest relatives, the monkeys and apes, 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. Humean thinking is making a major comeback.

Why did evolutionary biology stray from this path during the final quarter of the twentieth century? Why was morality considered unnatural, why were altruists depicted as hypocrites, and why were emotions left out of the debate? Why the calls to go against our own nature and to distrust a "Darwinian world"? The answer lies in what I have called the *Beethoven error*. In the same way that Ludwig van Beethoven

is said to have produced his beautiful, intricate compositions in one of the most disorderly and dirty apartments of Vienna, there is not much of a connection between the process of natural selection and its many products. The Beethoven error is to think that, since natural selection is a cruel, pitiless process of elimination, it can only have produced cruel and pitiless creatures (de Waal 2005).

But nature's pressure cooker does not work that way. It favors organisms that survive and reproduce, pure and simple. How they accomplish this is left open. Any organism that can do better by becoming either more or less aggressive than the rest, more or less cooperative, or more or less caring, will spread its genes.

The process does not specify the road to success. Natural selection has the capacity of producing an incredible range of organisms, from the most asocial and competitive to the kindest and gentlest. The same process may not have specified our moral rules and values, but it has provided us with the psychological makeup, tendencies, and abilities to develop a compass for life's choices that takes the interests of the entire community into account, which is the essence of human morality.

Appendix A

Anthropomorphism and Anthropodenial

Often, when human visitors walk up to the chimpanzees at the Yerkes Field Station, an adult female named Georgia (figure 8) hurries to the spigot to collect a mouthful of water before they arrive. She then casually mingles with the rest of the colony behind the mesh fence of their outdoor compound, and not even the best observer will notice anything unusual about her. If necessary, Georgia will wait minutes with closed lips until the visitors come near. Then there will be shrieks, laughs, jumps, and sometimes falls, when she suddenly sprays them.

This is not a mere "anecdote," as Georgia does this sort of thing predictably, and I have known quite a few other apes good at surprising naive people—and not just naive people. Hediger (1955), the great Swiss zoo biologist, recounts how even when he was fully prepared to meet the challenge, paying attention to the ape's every move, he nevertheless got drenched by an old chimpanzee with a lifetime of experience with this game.

Once, finding myself in a similar situation with Georgia

Figure 8 Georgia, our naughtiest chimpanzee, fascinated by her own reflection in the camera lens. Photograph by the author.

(i.e., aware that she had gone to the spigot and was sneaking up on me), I looked her straight in the eyes and pointed my finger at her warning, in Dutch, "I have seen you!" She immediately stepped away and let part of the water drop, swallowing the rest. I certainly do not wish to claim that she understands Dutch, but she must have sensed that I knew what she was up to, and that I was not going to be an easy target.

The curious situation in which scientists who work with these fascinating animals find themselves is that they cannot help but interpret many of their actions in human terms, which then automatically provokes the wrath of philosophers and other scientists, many of whom work with domestic rats, or pigeons, or with no animals at all. Unable to speak from firsthand experience, these critics must feel confident

indeed when they discard accounts by primatologists as anthropomorphie, and explain how anthropomorphism is to be avoided.

Although no reports of spontaneous ambush tactics in rats have come to my attention, these animals could conceivably be trained with patient reinforcement to retain water in their mouth and stand amongst other rats. And if rats can learn to do so, what is the big deal? The message of the critics of anthropomorphism is something along the lines of "Georgia has no plan; Georgia does not know that she is tricking people; Georgia just learns things faster than a rat." Thus, instead of seeking the origin of Georgia's actions within herself, and attributing intentions to her, they propose to seek the origin in the environment and the way it shapes behavior. Rather than being the designer of her own disagreeable greeting ceremony, this ape fell victim to the irresistible rewards of human surprise and annoyance. Georgia is innocent!

But why let her off the hook that easily? Why would any human being who acts this way be scolded, arrested, or held accountable, whereas any animal, even of a species that resembles us so closely, is considered a mere passive instrument of stimulus-response contingencies? Inasmuch as the absence of intentionality is as difficult to prove as its presence, and inasmuch as no one has ever proven that animals differ fundamentally from people in this regard, it is hard to see the scientific basis for such contrasting assumptions. Surely, the origin of this dualism is to be found partly outside of science.

The dilemma faced by behavioral science today can be summarized as a choice between cognitive and evolutionary parsimony (de Waal 1991, 1999). *Cognitive parsimony* is the

traditional canon of American Behaviorism. It tells us not to invoke higher mental capacities if we can explain a phenomenon with ones lower on the scale. This favors a simple explanation, such as conditioned behavior, over a more complex one, such as intentional deception. This sounds fair enough (but see Sober 1990). *Evolutionary parsimony*, on the other hand, considers shared phylogeny. It posits that if closely related species act the same, the underlying mental processes are probably the same, too. The alternative would be to assume the evolution of divergent processes that produce similar behavior, which seems a wildly uneconomic assumption for organisms with only a few million years of separate evolution. If we normally do not propose different causes for the same behavior in, say, dogs and wolves, why should we do so for humans and chimpanzees?

In short, the cherished principle of parsimony has taken on two faces. At the same time that we are supposed to favor low-level over high-level cognitive explanations, we also should not create a double standard according to which shared human and ape behavior is explained differently. If accounts of human behavior commonly invoke complex cognitive abilities—and they most certainly do (Michel 1991)—we must carefully consider whether these abilities are perhaps also present in apes. We do not need to jump to conclusions, but the possibility should at least be allowed on the table.

Even if the need for this intellectual breathing room is most urgently felt in relation to our primate relatives, it is neither limited to this taxonomic group nor to instances of complex cognition. Students of animal behavior are faced with a choice between classifying animals as automatons or granting them volition and information-processing capacities. Whereas one school warns against assuming things we cannot prove,

another school warns against leaving out what may be there: even insects and fish come across to the human observer as internally driven, seeking, wanting systems with awareness of their surroundings. Descriptions that place animals closer to us than to machines adopt a language that we customarily use for human action. Inevitably, these descriptions sound anthropomorphic.

Obviously, if anthropomorphism is defined as the misattribution of human qualities to animals, no one wishes to be associated with it. But much of the time, a broader definition is employed, namely the description of animal behavior in human, hence intentionalistic, terms. Even though no anthropomorphism proponent would propose to apply such language uncritically, even the staunchest opponents of anthropomorphism do not deny its value as a heuristic tool. It is this use of anthropomorphism as a means to get at the truth, rather than as an end in itself, that distinguishes its use in science from that by the layperson. The ultimate goal of the anthropomorphizing scientist is emphatically not the most satisfactory projection of human feelings onto the animal, but testable ideas and replicable observations.

This requires great familiarity with the natural history and special traits of the species under investigation, and an effort to suppress the questionable assumption that animals feel and think like us. Someone who cannot imagine that ants taste good cannot successfully anthropomorphize the anteater. So, in order to have any heuristic value at all, our language must respect the peculiarities of a species while framing them in a way that strikes a chord in the human experience. Again, this is easier to achieve with animals close to us than with animals, such as dolphins or bats, that move through a different medium or perceive the world through

different sensory systems. Appreciation of the diversity of *Umwelten* (von Uexküll 1909) in the animal kingdom remains one of the major challenges of the student of animal behavior.

The debate about the use and abuse of anthropomorphism, which for years was confined to a small academic circle, has recently been thrust into the spotlight by two books: Kennedy's (1992) *The New Anthropomorphism*, and Marshall Thomas's (1993) *The Hidden Life of Dogs*. Kennedy reiterates the dangers and pitfalls of assuming higher cognitive capacities than can be proven, thus defending cognitive parsimony. Marshall Thomas, on the other hand, does not make any bones about the anthropomorphic bias of her informal study of canine behavior. In her best-seller, the anthropologist lets virgin bitches "save" their virginity for future "husbands" (i.e., ignore sexual attentions prior to meeting a favorite male, p. 56), watches wolves set out for the hunt without "pitying themselves" (p. 39), and looks into her dogs' eyes during a vicious gang attack seeing "no anger, no fear, no threat, no show of aggression, just clarity and overwhelming determination" (p. 68).

There is quite a difference between the use of anthropomorphism for communicatory purposes or in order to generate hypotheses, and the sort of anthropomorphism that does little else than project human emotions and intentions onto animals without justification, explication, or investigation (Mitchell et al. 1997). The uncritical anthropomorphism of Marshall Thomas is precisely what has given the practice a bad name, and has led critics to oppose it in all of its forms and disguises. Rather than let them throw out the baby with the bathwater, however, the only question that needs to be answered is whether a certain dose of anthropomorphism,

used in a critical fashion, helps or hurts the study of animal behavior. Is it something that, as Hebb (1946) already noted, allows us to make sense of animal behavior, and, as Cheney and Seyfarth (1990: 303) declared, "works" in that it increases the predictability of behavior? Or is it something that, as Kennedy (1992) and others argue, needs to be brought under control, almost like a disease, because it makes animals into humans?

While it is true that animals are not humans, it is equally true that humans are animals. Resistance to this simple yet undeniable truth is what underlies the resistance to anthropomorphism. I have characterized this resistance as *anthropodenial*, the *a priori* rejection of shared characteristics between humans and animals. Anthropodenial denotes willful blindness to the human-like characteristics of animals, or the animal-like characteristics of ourselves (de Waal 1999). It reflects a pre-Darwinian antipathy to the profound similarities between human and animal behavior (e.g., maternal care, sexual behavior, power seeking) noticed by anyone with an open mind.

The idea that these similarities require unitary explanations is anything but new. One of the first to advocate cross-specific explanatory uniformity was David Hume (1985 [1739]: 226), who formulated the following touchstone in *A Treatise of Human Nature*:

> Tis from the resemblance of the external actions of animals to those we ourselves perform, that we judge their internal likewise to resemble ours; and the same principle of reasoning, carry'd one step farther, will make us conclude that since our internal actions resemble each other, the causes, from which they are deriv'd, must also be resembling. When

> any hypothesis, therefore, is advanc'd to explain a mental operation, which is common to men and beasts, we must apply the same hypothesis to both.

It is important to add that, in contrast to American behaviorists, who two centuries after Hume would accommodate animals and humans within a single framework by seriously downgrading human mental complexity and relegating consciousness to the domain of superstition (e.g., Watson 1930), Hume (1985 [1739]: 226) held animals in high esteem, writing that "no truth appears to me more evident than that beasts are endow'd with thought and reason as well as men."

Strictly speaking, one cannot boast a unified theory of all behavior, human and animal, while at the same time decrying anthropomorphism. After all, anthropomorphism assumes similar experiences in humans and animals, which is exactly what one would expect in case of shared underlying processes. The behaviorists' opposition to anthropomorphism probably came about because no sane person would take seriously their claim that internal mental operations in *our* species are a figment of the imagination. The masses refused to accept that their own behavior could be explained without considering thoughts, feelings, and intentions. Don't we have mental lives, don't we look into the future, aren't we rational beings? Eventually, the behaviorists relented, exempting the bipedal ape from their theory of everything.

This is where the problem for other animals began. Once cognitive complexity was admitted in humans, the rest of the animal kingdom became the sole light-bearer of Behaviorism. Animals were expected to follow the law of effect to the absolute letter, and anyone who thought differently was just being anthropomorphic. Attribution of human-like experiences

to animals was declared a cardinal sin. From a unified science, Behaviorism had deteriorated into a dichotomous one with two separate languages: one for human behavior, another one for animal behavior.

So, the answer to the question "Isn't anthropomorphism dangerous?" is that, yes, it is dangerous to those who wish to uphold a wall between humans and other animals. It places all animals, including humans, on the same explanatory plane. It is hardly dangerous, though, to those working from an evolutionarily perspective so long as they treat anthropomorphic explanations as hypotheses (Burghardt 1985). Anthropomorphism is a possibility among many, but one to be taken seriously given that it applies intuitions about ourselves to creatures very much like us. It is the application of human self-knowledge to animal behavior. What could be wrong with that? We apply human intuition in mathematics and chemistry, so why should we suppress it in the study of animal behavior? Stronger yet: does anyone truly believe that anthropomorphism is avoidable (Cenami Spada 1997)?

In the end we must ask: What kind of risk are we willing to take, the risk of underestimating animal mental life or the risk of overestimating it? There is a symmetry between anthropomorphism and anthropodenial, and since each has its strengths and weaknesses, there is no simple answer. But from an evolutionary perspective, Georgia's mischief is most parsimoniously explained in the same way we explain our own behavior—as the result of a complex, and familiar, inner life.