A Review of *Primates and Philosophers: How Morality Evolved*  

By Andrew McAninch**


** Doctoral student in philosophy at Indiana University, Bloomington.

It is hard to deny the importance of Frans de Waal’s work on primate behavior, which spans over three decades, including the last 18 years at the Yerkes National Primate Research Center at Emory University. De Waal’s work has contributed to a substantial shift in scientific and popular thinking about animal social behavior. Specifically, de Waal’s research on primate peacemaking, conflict resolution, and pro-social behavior has discredited the assumption that primates are purely selfish, individualistic, and aggressive beasts. His work makes the case that not only do many primates exhibit truly other-regarding behavior but also that this behavior is as deeply rooted in their nature as is any instinct for self-preservation. Primates behave pro-socially not despite their biological dispositions but rather as a natural manifestation of them. In short—de Waal has argued—primates, our closest evolutionary ancestors, are fundamentally good-natured.

In the recent volume *Primates and Philosophers: How Morality Evolved*, edited by Josiah Ober and Stephen Macedo, and based on de Waal’s 2003 Tanner Lectures on Human Values, he extends that thesis into an exploration of the evolutionary origins of human morality. De Waal’s central claim in the lead essay is that the pro-social emotional dispositions of non-human primates constitute the “building blocks” of human morality. Following de Waal’s essay are separate responses by journalist and science writer Robert Wright, philosopher of science Philip Kitcher, and moral philosophers Christine M. Korsgaard and Peter Singer. De Waal concludes the volume with a response to his commentators.

*Primates and Philosophers* demonstrates some of the perils of addressing a topic of so much importance to so many from perspectives as varied as evolutionary biology, experimental psychology, and moral philosophy. I will discuss some of these perils below. But, more importantly, let me emphasize that this volume also demonstrates the promise, benefit, and, indeed, inevitability of bringing these perspectives into the same conversation. I recommend this volume to anyone interested in the evolutionary origins of morality. *Primates and Philosophers* is rich enough to interest scientists and philosophers, but it is also accessible
enough to engage students who are new to the topic. Indeed, it could easily be used as an instructional tool in an undergraduate course on evolution and ethics.

After canvassing the central claims and arguments of de Waal’s lead essay, I will discuss some of the main concerns that arise between de Waal and his commentators.

The Evolution of Morality and the Rejection of Veneer Theory

Discourse about the “evolution of morality” is ambiguous. “Morality,” in one sense, could simply refer to a set of moral rules or principles (e.g., the Golden Rule, the Principle of Utility). But “morality,” in another sense, could refer to the set of human practices in which such moral principles are implicated—actions that are guided by such principles, or judged in terms of them, or otherwise mediated by moral beliefs or attitudes. It is this latter phenomenon that de Waal is investigating. He is interested in those underlying psychological capacities without which an individual’s behavior could not qualify as either moral or immoral (16). I will call this set of relevant capacities “moral agency.” Thus, de Waal’s guiding question is whether the foundations of human moral agency are to be found in the psychological and behavioral dispositions of our evolutionary ancestors. And his answer is a resounding “Yes.” The “building blocks” of human moral agency, he argues, are present in other primates, and they are constituted by a range of other-regarding emotions and behavioral responses whose evolutionary function was to facilitate social coordination and cohesion.

De Waal bases his conclusion on both methodological considerations and a rich supply of empirical observations. Methodologically, de Waal pits his argument against what he calls the “Veneer Theory” of morality, a theory he attributes to T.H. Huxley (an attribution disputed by Kitcher), G.C. Williams, Richard Dawkins, and Robert Wright (an attribution disputed by Wright). Veneer Theory, as de Waal characterizes it, says little about what constitutes moral agency; rather, it concerns whether or not the capacities underlying our apparent moral agency are natural, and it holds that they are not. Veneer Theory, de Waal writes,

...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. ... [Veneer Theory] 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 (6).

Because, according to this gloss, Veneer Theory holds that “deep down we are not truly moral,” it seems to hold that moral agency is, strictly speaking, illusory. Huxley, however, affirms that we can exercise genuine moral agency but that we do so despite our nature—we are moral gardeners who must constantly keep the weeds of our nature at bay (7). The biologist Michael Ghiselin, though, suggests that, if we are not moral by nature, then we are not really moral at all. What
appears to be morality is just that—an appearance, a veneer—a thesis encapsulated in his famous words, “Scratch an ‘altruist,’ and watch a ‘hypocrite’ bleed” (10).¹

Veneer Theory faces a number of problems, according to de Waal. First, because Veneer Theorists deny that our moral tendencies are natural, they appear to be at a loss for how to explain scientifically why we have these tendencies and how we exercise them. Second, Veneer Theory’s assumption that we are fundamentally asocial and selfish—and that the “social contract” developed only in service of individual self-interest—flies in the face of overwhelming empirical evidence to the contrary. De Waal writes, “[T]here 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” (4). The state of nature may have been “nasty, brutish, and short”; but for humans and their primate ancestors, it has always been “social to the core” (5). Third, if de Waal is right that our primate ancestors demonstrate a range of other-regarding behavioral dispositions, then the most evolutionarily parsimonious explanation of human other-regarding dispositions is that they are homologous with our ancestors’ traits.

De Waal supports his positive thesis by elaborating on the second and third points above and by citing empirical work on animal behavior collected over decades of research. The observations he shares about primate social behavior are fascinating, and his findings are at once surprising and plausible. For example, de Waal writes about Krom, a chimpanzee at the Arnhem Zoo, who one day became interested in a water-filled tire hanging from a log. Her attempts to extract the tire merely put it further out of reach. Witnessing Krom’s struggle, Krom’s seven-year-old “nephew” Jakie intervened to help, pushing away the other tires and gently lowering the water-filled tire. He placed it in front of Krom, who proceeded to drink up the water with relish (32).

Another example is consolation behavior in chimpanzees. After an aggressive conflict between two chimps, a bystander will often console the “loser” of the conflict, by putting an arm around him for example. In such cases of consolation, there is no clear benefit to the consoling party (33-4).

De Waal has also observed that chimpanzees are more likely to share food with chimps that have groomed them earlier, thus expressing a kind of gratitude (43).

These observations, especially the first two, purport to reveal a concern for others that is neither accidental nor reducible to indirect self-interest. Rather, de Waal argues, these behaviors represent the exercise of a capacity to empathize with other individuals, a capacity whose core is a simple Perception-Action Mechanism (PAM). The PAM, explains de Waal, “provides an observer (the ‘subject’) with access to the emotional state of another (the ‘object’) through the subject’s own neural and bodily representations” (37). He continues,
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 (37).

The existence of the PAM itself is evolutionarily explicable; such a mechanism would have conferred a significant evolutionary benefit to our primate ancestors, whose individual survival depended on cooperation and coordination with conspecifics. And, yet, this very mechanism would have also opened the door to a concern for others that transcends self-preservation—i.e., moral concern (15).

If empathy is what makes moral agency possible, and, if empathy evolved through the natural selection of the PAM, then the roots of moral agency are indeed natural, contra Veneer Theory.

De Waal’s commentators all reject the simplistic version of Veneer Theory that he sets up as his foil. And, although some of them (Kitcher, for example) take issue with de Waal’s conclusions about the extent of primates’ altruistic tendencies, all of them agree that de Waal has revealed a non-trivial level of empathic response in non-human primates. Where they disagree with de Waal is on the question of whether empathy and other “moral sentiments” constitute the foundations of human moral agency. At issue, then, is whether these emotional dispositions are the fundamental capacities underlying full-blooded moral agency. This issue proves to be paramount in the commentators’ responses to de Waal, to which I will turn now.

Cognitive Mechanisms and Emotional Mechanisms

Robert Wright’s main objection to de Waal concerns the kind of anthropomorphic language that de Waal sometimes uses. Wright claims that de Waal slides between emotional and cognitive anthropomorphic language. That is, he sometimes explains an animal’s behavior in terms of its guidance by emotions (empathy, personal distress, anger) and sometimes in terms of strategic calculation or conscious deliberation (reasoning, intending, deciding). In most cases, both types of explanation are compatible with the observed behavior—and after all, many of our emotional dispositions are “proxies for strategic calculation” in the sense that they were naturally selected for their benefit in efficiently achieving some goal of the organism (86). But, Wright argues, in the absence of independent evidence that non-human primates possess the capacities for strategic calculation and conscious deliberation—linguistic ability, for example—de Waal should restrict his language to emotional anthropomorphism.
Wright is less clear about what is at stake, with respect to moral agency, between positing conscious deliberation and positing the exercise of involuntary emotional mechanisms. If the latter are “proxies” for the former, then is there anything significant that they lack but that the former has? Or are they simply two routes to the same goal? Korsgaard and Kitcher suggest an answer to these questions.

**Wantons and Persons**

Korsgaard and Kitcher invoke the philosopher Harry G. Frankfurt’s distinction between a *wanton* and a *person.*

A wanton is a creature that acts on whatever desire or impulse is strongest at that moment, without concern for whether or not that desire or impulse *ought* to be effective in moving the creature. Korsgaard and Kitcher both argue that de Waal has shown, at most, that non-human primates are empathic wantons. Primates’ actions are, on occasion, motivated by involuntary, empathic emotional impulses. (Kitcher includes a careful discussion of the degrees of psychological altruism and agrees with de Waal that primates have demonstrated altruism of “moderate intensity” (130).) But because they lack the ability to step back from and reflect on these impulses—and to be guided by their reflective judgments about their motivations—they fall short of being persons. They fail to satisfy a necessary condition for moral agency.

Korsgaard expresses this idea in Kantian terms. Although non-human primates act from truly other-regarding motivations, they are essentially “pushed and pulled” by these impulses. They lack the *autonomy* with which a person can evaluate those impulses and thus choose which ends she will pursue. Korsgaard writes, “The morality of your action is not a function of the content of your intentions [or emotions]. It is a function of the exercise of normative self-government” (112). Again, at issue between de Waal and his commentators is not whether non-human primates demonstrate empathy; rather, it is whether empathy constitutes the fundamental criterion of moral agency. Although Korsgaard associates the point about reflective self-government with Kant, Kitcher finds parallel considerations in Hume and Adam Smith (132-3). And even Peter Singer, a utilitarian, registers agreement with Kant that the capacity to reflect on our given emotional responses is central to moral agency (149-50).

**The Scope of Moral Concern**

One more issue that arises in the commentaries bears mentioning. De Waal notes in his essay that other-regarding emotional dispositions evolved as a distinctly in-group phenomenon and that this in-group/out-group mentality exists today in our current moral practices (53). The point here is that moral concern, although by definition extending beyond oneself, only rarely and tenuously extends beyond one’s own community. It is crucial to understand this claim as a descriptive observation about how people are in fact inclined to regard those
outside their community and not as a *normative* claim about the appropriate scope of moral principles. De Waal sometimes seems to elide this distinction.

Singer, for example, argues that, because the other-regarding emotional dispositions cited by de Waal are always limited in scope, they fail to demonstrate the feature of impartiality that is definitive of moral concern as a normative principle (145). De Waal’s response to Singer is somewhat puzzling. He agrees that primates’ other-regarding concern is limited to their *in-group*, but he appears to take the further step of *endorsing* this limited scope of concern. “It is not just that we are biased in favor of the innermost circles (ourselves, our family, our community, our species),” he writes, “we *ought* to be. Loyalty is a moral duty” (165). Now, to make this claim is to take a substantive moral position, and, as such, it requires a moral argument. The apparent conflict between moral impartiality and personal loyalty is an area of considered philosophical debate. But to reach de Waal’s conclusion, it is not sufficient to say that, as a matter of fact, most of us demonstrate in-group biases. It is also not sufficient to argue—as de Waal seems to do—that in-group biases are morally valuable in virtue of their *instrumental* value to group cohesion. This argument begs the question: after all, the unargued privilege of one’s own group over others is precisely what the notion of moral impartiality throws into question.

* * *

The debate between de Waal and his commentators reflects both the importance and the difficulty of addressing these two questions together:

(1) Does human moral agency have its roots in our evolutionary ancestors?
(2) What are the criteria of human moral agency anyway?

De Waal’s commentators all argue, either implicitly or explicitly, that a developed answer to the second question throws some doubt on de Waal’s positive answer to the first question. Kitcher, for example, regards de Waal’s talk about the “building blocks” of morality to be unacceptably vague—the problem being that, although de Waal has garnered impressive evidence for pro-social tendencies in primates, he “hasn’t thought as hard about the human phenomenon he takes to be anticipated or foreshadowed in chimpanzee social life” (123). The commentators argue convincingly that the criteria of human moral agency include far more than empathic or altruistic dispositions. Of course, none of this is to deny that de Waal’s emphasis on the *continuities* between humans and other primates broadens the scope of inquiry in which moral philosophers and psychologists must work. And that is a notable achievement in itself.
De Waal is fond of pointing out the prevalence of the “Beethoven Error” in much evolutionary thinking: one might assume that a “cruel, pitiless process of elimination” can only produce “cruel and pitiless creatures” (58). But, to the contrary, just as Beethoven wrote some his most beautiful music in one of the filthiest apartments in Vienna, “Mother Nature” produced some of humanity’s most gentle and virtuous dispositions through the harsh process of natural selection.

De Waal spends a considerable amount of time elaborating on the complexities of empathic tendencies in primates and other animals. On one end of the spectrum is mere **emotional contagion**—the inducement, in a subject, of an emotional state similar to that perceived in the object—and consequent **personal distress** in the subject (26). On the other end of the spectrum is **cognitive empathy**, in which a subject’s empathic response is mediated by his awareness of the object’s specific predicament, including, perhaps, how this predicament appears from the object’s point of view. Cognitive empathy seems to require a capacity for self-awareness that “allows the other’s situation to be divorced from one’s own while maintaining the emotional link that motivates behavior” (36). It is cognitive empathy that is implicated in acts of **targeted helping**—such as Jakie’s act of helping Krom—acts that most plausibly appear altruistic and morally meaningful.

In an appendix preceding Wright’s response, de Waal actually addresses this issue. He accepts that there is a tension between **cognitive parsimony**, the principle that we should assume lower-level mental capacities whenever possible, and **evolutionary parsimony**, the principle that we should assume the same mental capacities are at work in closely related species (62).