

Headwaters

Volume 25

Article 11

2008

Natural Selection and Moral Sentiment: Evolutionary Biology's Challenge to Moral Philosophy

Charles W. Wright

College of Saint Benedict/Saint John's University, cwright@csbsju.edu

Follow this and additional works at: <https://digitalcommons.csbsju.edu/headwaters>



Part of the [Biology Commons](#), and the [Ethics and Political Philosophy Commons](#)

Recommended Citation

Wright, Charles W. (2008) "Natural Selection and Moral Sentiment: Evolutionary Biology's Challenge to Moral Philosophy," *Headwaters: Vol. 25*, 115-120.

Available at: <https://digitalcommons.csbsju.edu/headwaters/vol25/iss1/11>

This Article is brought to you for free and open access by DigitalCommons@CSB/SJU. It has been accepted for inclusion in Headwaters by an authorized editor of DigitalCommons@CSB/SJU. For more information, please contact digitalcommons@csbsju.edu.

Natural Selection and Moral Sentiment: Evolutionary Biology's Challenge to Moral Philosophy

It's unfortunate that our college curricula don't do a better job of enabling students to see the interconnections among the various disciplines. It's regrettable that faculty in different disciplines find it so hard to talk to each other about the things that interest them the most. It's too bad we have so much trouble overcoming the "over" in "over-specialization."

It was just over 30 years ago that E. O. Wilson had the temerity to suggest that ethics ought one day to become a sub-discipline of biology. This arrangement would, he said, foster greater progress than had heretofore been achieved because, in his memorable words, moral philosophers had so far done little more than consult "the emotive centers of their own hypothalamic–limbic system" (Wilson 1975, 563). Needless to say, philosophers were not particularly impressed by his proposal. But with the benefit of 30 years' hindsight, Wilson's bold assessment begins to look prescient — wrong in the details, to be sure (nobody supposes now that moral emotions might be located in the hypothalamus), but right in principle (emotionally structured moral intuitions are now thought to play a central role in everyday moral judgment).

In defense of this claim, I shall first briefly review the evolutionary biological background to Wilson's pronouncement and take note of one of the sentimentalist implications arising from it. Then I'll review contemporary research by social psychologists and neuroscientists that provides empirical warrant for these theoretical implications. I'll conclude by considering briefly what these developments might mean for the future of moral philosophy.

Popular conceptions of Darwinian evolutionary theory have long supposed that natural selection only fosters ruthless competition among organisms. But at the same time such social Darwinian ideas circulated in the public sphere, biologists had come to the conclusion that selection also favors cooperative behavior. Starting with Darwin himself, until the mid-20th century, leading scientists explained cooperative behavior among animals using models of group selection which proposed that such behavior

evolved because of the benefits it conferred on the group, be it the local population or the species as a whole.

Such proposals were thoroughly debunked, however, with the publication of George Williams' *Adaptation and Natural Selection* in 1966. After Williams' critique the only reputable scientific position was to suppose that the forces of natural selection worked on the individual organism (or, on the genetic material that it contained), not on groups. But then "altruism" became a problem in need of a solution. Certain kinds of cooperative behavior — alarm calls, for instance — no longer made theoretical sense, since they impose fitness-reducing costs upon one organism while conferring fitness-enhancing benefits to others. From a narrow Darwinian perspective — the perspective of popular imagination — such behavior ought to disappear from a population.

The theoretical advances made in response to this problem are by now fairly widely known. W. D. Hamilton (1964) proposed the idea of inclusive fitness, whereby one organism's apparently fitness-reducing behavior will ensure the propagation of its genetic material when directed toward sufficiently close relatives. Robert Trivers (1971) next hypothesized that under certain conditions — for instance, high frequency of interaction, relatively long lifespan, a certain cost-benefit ratio to the acts, and so forth — reciprocal exchanges of benefits between unrelated individuals could develop. R. D. Alexander's (1979, 1987) concept of indirect reciprocity extended Trivers' model by including social observation and reputation formation, factors enabling stable networks of exchange to develop among relatively small groups of individuals. Among these three theoretical approaches, cooperative behavior could be explained among close kin, friends and allies, and small face-to-face communities. One philosophically interesting feature of these theoretical models is their hypothesis that the regulation of these cooperative relations would be supported by emotional, rather than cognitive systems. This was the point of Wilson's offhand remark about philosophers consulting their hypothalamic systems.

Why this hypothesis? Explanations differ, but all point to some selective advantage that emotional regulation would have over cognitive regulation of behavior. Trivers (1971) suggests that motives embedded in affective systems would better sustain stable networks of reciprocity than those produced by rational calculations of advantage. Michael Ruse (1986, 1988), an early philosophical convert to the evolutionary perspective, appeals to energetic parsimony. Emotional processing systems — automatic, unconscious, and carrying strong action dispositions — are, in terms of the time, energy, and attention required to influence behavior, both cheaper to support and more efficient than cognitive processing systems. And natural selection always favors cheap efficiency. Whatever the hypothesized reason, the perspective endorsed by these

theorists more or less amounted, in the case of human social conduct, to an endorsement of moral sentimentalism.

Unsurprisingly, moral philosophers have contested the adequacy of such accounts. Patricia Williams (1993) insisted that evolutionary models of the moral sentiments entail an unwarranted constriction on human freedom of choice. Alan Gewirth (1993) argued that such models reduce morality to a kind of tribalism and utterly fail to account for the universal and egalitarian moral principles that inform modern Western societies. And Philip Kitcher (1985) claimed that Wilson's sentimentalist proposal committed him to naïve conceptions of emotivism and relativism, against which John Rawls' constructivist account of moral objectivity was much to be preferred. But even as philosophical reason debunked evolutionary models of moral sentiments, recent empirical research suggests that the biologists cannot be so easily dismissed.

Space constraints won't permit me to review this literature in the detail it warrants. I can offer you only the broadest of brushstrokes. There is, to start, the widely cited work of neuroscientist Antonio Damasio (1994) and his colleagues (Anderson, Bechara, Damasio, Tranel, & Damasio, 1999; Bechara, Damasio, & Damasio, 2000; Bechara, Tranel, Damasio, & Damasio, 1996), who found that damage to certain emotional centers in the brain severely impede a person's capacity for moral judgment. Similarly, social psychologist R. J. R. Blair (1995, 1997; Blair, Mitchell, Kelly, Richell, Leonard, Newman, & Scott, 2002) has found that the behavioral disorders of psychopathic individuals are better explained in terms of emotional deficits rather than in terms of shortcomings in their capacity for reason. Neuroscientist Jorg Moll and his colleagues (Moll, Eslinger, & de Oliveira-Souza, 2001; Moll, de Oliveira-Souza, Bramati, & Grafman 2002; Moll, de Oliveira-Souza, & Eslinger, 2003), who study the functional architecture of brains engaged in moral evaluation, also find that moral judgments elicit greatest activity in brain regions associated with affect.

From the perspective of contemporary moral philosophy, though, the most interesting work may have been carried out by Joshua Greene, R. Brian Sommerville, Leigh Nystrom, John Darley, and Jonathan Cohen, who asked subjects to deliberate about a much discussed moral problem while undergoing fMRI scans of their brain activity. In the problem, a runaway trolley will kill five people unless some action is taken. The trolley might be switched to another track where only one person would be killed, or a large person might be pushed in front of the trolley to bring it to a halt. Greene and his coworkers (2001) found that depending on the nature of the proposed action — switching versus pushing — different regions of the brain tended to be differentially activated. The “personal” dilemma — push a person in front of the trolley — energized brain areas associated with social emotional processing more highly. By contrast,

regions associated with classically cognitive processes such as working memory, abstract reasoning, and problem solving were more highly activated by the “impersonal” option of rerouting the trolley away from the five and toward one.

In a follow-up study, Greene and his colleagues (Greene, Nystrom, Engell, Darley, & Cohen, 2004) investigated the brain processes of subjects who deliberate about and in some cases endorse actions (pushing a person in front of a trolley, smothering a baby) that would under ordinary circumstances be judged morally repugnant, but because of morally salient circumstances (preventing the deaths of even more people) might be judged acceptable. These difficult personal moral dilemmas elicited activity not only in emotional and cognitive centers, but also in a region — the anterior cingulate cortex — recently associated with the mediation of cognitive conflict. By contrast, in cases of straightforward personal transgressions that lack countervailing moral considerations — for example, a teenage mother who kills an unwanted infant — social-emotional centers are significantly more highly energized.

Greene et al. (2004) interpret their findings in the first instance as evidence for “a synthetic view of moral judgment that acknowledges the crucial roles played by both emotion and ‘cognition’.” (397) But they also situate this dual-process model in an evolutionary perspective. Evidence from the social dynamics of primates (De Waal 1996), from the archaeological record (Aiello & Dunbar 1993; Dunbar 1996), as well as from existing foraging societies (Boehm 1999) all support the hypothesis that humankind’s early hominid ancestors lived in tightly knit social networks in which behavior was regulated by moral emotions such as a guilt, anger, empathy, gratitude, and a sense of fairness. On the basis of this evidence they conclude that “it would be strange if human behavior were not driven in part by domain-specific social-emotional dispositions.” (Greene et al., 2004, 389) By contrast, the “domain general capacity for sophisticated abstract reasoning” (390) found to be at work in deliberations about impersonal and difficult personal moral dilemmas is a relatively more recent evolutionary development. The clear implication is that phylogenetically older affective systems steer everyday moral conduct while more recently evolved cognitive systems are only activated in situations of conflict or heightened social complexity.

What are the implications of this rapidly growing body of research for the future of moral philosophy? The first, I think, is that moral philosophers ignore the evolutionary history of the capacity for moral judgment at the risk of disciplinary irrelevance. Debates between rationalists and sentimentalists, for instance, are doomed to sterility until both sides recognize that evolution has constructed moral judgment out of both affective and cognitive components. In certain circumstances judgments will conform to the sentimentalist model, while other judgments in other circumstances will con-

form more closely to the rationalist model. No judgment will ever be devoid of either affect or cognition. Philosophical work built on premises contrary to these neurological facts may advance academic careers but seems unlikely to produce genuine insight.

Similarly, the question whether moral judgments are necessarily or contingently motivating is unlikely to reach settlement until philosophers recognize that moral judgment is not a unitary phenomenon, but rather arises from the coordinated activity of a variety of brain sectors (Greene & Haidt, 2002; Moll et al., 2003). Given this neurological structure, both the internalist and the externalist accounts can be correct depending on the psychological system giving rise to a judgment. If a judgment is produced by automatic, intuitive systems, the internalist account will be closer to the truth. If produced by controlled cognitive processes, externalist descriptions will be more appropriate. In those cases where a moral judgment involves both automatic and controlled systems of cognition, both accounts will be correct.

A full appreciation of the evolutionary history and multi-layered structure of moral judgment promises dramatically to reshape the philosophical conversation about judgment and motivation. The second implication is more general, but follows from the preceding observations. It would seem that it is time for philosophers to relinquish the conceit that moral philosophy — whether in Kantian, conceptual analytic, or any other form — can be an autonomous, *a priori* undertaking.

Chuck Wright is an Associate Professor of Philosophy.

Works Cited

- Aiello, Leslie C. & R. I. M. Dunbar (1993). "Neocortex Size, Group Size, and the Evolution of Language," *Current Anthropology* 34(23), April, 184–193.
- Alexander, Richard D. (1979). *Darwinism and Human Affairs* (Seattle, WA: University of Washington Press).
- Alexander, Richard D. (1987). *The Biology of Moral Systems* (Hawthorne, NY: A. de Gruyter).
- Anderson, Steven W., Antoine Bechara, Hanna Damasio, Daniel Tranel, & Antonio R. Damasio (1999). "Impairment of Social and Moral Behavior Related to Early Damage in Human Prefrontal Cortex," *Nature Neuroscience* 2(11), November, 1032–1037.
- Bechara, Antoine, Hanna Damasio, & Antonio R. Damasio (2000). "Emotion, Decision Making and the Orbitofrontal Cortex," *Cerebral Cortex* 10, March, 295–307.
- Bechara, Antoine, Daniel Tranel, Hanna Damasio, & Antonio Damasio (1996). "Failure to Respond Autonomically to Anticipated Future Outcomes Following Damage to Prefrontal Cortex," *Cerebral Cortex* 6, March/April, 215–225.
- Blair, R. J. R. (1995). "A Cognitive Developmental Approach to Morality: Investigating the Psychopath," *Cognition* 57, 1–29.

- Blair, R. J. R. (1997). "Moral Reasoning and the Child with Psychopathic Tendencies," *Personality and Individual Differences* 22(5), 731–739.
- Blair, R. J. R., D. G. V. Mitchell, R. A. Richell, S. Kelly, A. Leonard, C. Newman, & S. K. Scott (2002). "Turning a Deaf Ear to Fear: Impaired Recognition of Vocal Affect in Psychopathic Individuals," *Journal of Abnormal Psychology* 111(4), 682–686.
- Boehm, Christopher (1999). *Hierarchy in the Forest: The Evolution of Egalitarian Behavior* (Cambridge, MA: Harvard University Press).
- Damasio, Antonio (1994). *Descartes' Error* (New York: HarperCollins).
- De Waal, Frans (1996). *Good Natured: The Origins of Right and Wrong in Humans and Other Animals* (Cambridge, MA: Harvard University Press).
- Dunbar, Robin (1996). *Grooming, Gossip, and the Evolution of Language* (Cambridge, MA: Harvard University Press).
- Gewirth, Alan (1993). "How Ethical is Evolutionary Ethics," in M. Nitecki & D. Nitecki (Eds.), *Evolutionary Ethics* (Albany, NY: SUNY Press), pp. 241–256.
- Greene, Joshua D. & Jonathan Haidt (2002). "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences* 6(12), December, 517–523.
- Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darley, & Jonathan D. Cohen (2004). "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron* 44, October 14, 389–400.
- Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, & Jonathan D. Cohen (2001). "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293, September 14, 2105–2108.
- Kitcher, Philip (1985). *Vaulting Ambition: Sociobiology and the Quest for Human Nature* (Cambridge, MA: MIT Press).
- Moll, Jorg, Paul J. Eslinger, & Ricardo de Oliveira-Souza (2001). "Frontopolar and Anterior Temporal Cortex Activation in a Moral Judgment Task," *Arquivos de Neuro-Psiquiatria* 59(3-B), 657–664.
- Moll, Jorg, Ricardo de Oliveira-Souza, Ivelene E. Bramati, & Jordan Grafman (2002). "Functional Networks in Emotional Moral and Nonmoral Social Judgments," *Neuroimage* 16, 696–703.
- Moll, Jorg, Ricardo de Oliveira-Souza, & Paul J. Eslinger (2003). "Morals and the Human Brain: A Working Model," *NeuroReport* 14(3), March, 299–304.
- Ruse, Michael (1986). "Evolutionary Ethics: A Phoenix Arisen," *Zygon* 21, 95–112.
- Ruse, Michael (1988). "Evolutionary Ethics: Healthy Prospect or Last Infirmity," *Canadian Journal of Philosophy*, Supplementary Volume 14, *Philosophy & Biology*, M. Matthen & B. Linsky (Eds.) (Calgary, Alberta, Canada: University of Calgary Press), 27–71.
- Trivers, Robert (1971). "The Evolution of Reciprocal Altruism," *Quarterly Review of Biology* 46, March, 35–57.
- Williams, George (1966/1996). *Adaptation and Natural Selection* (Princeton, NJ: Princeton U. Press).
- Williams, Patricia A. (1993). "Can Being Whose Ethics Evolved Be Ethical Beings," in M. Nitecki & D. Nitecki (Eds.), *Evolutionary Ethics* (Albany, NY: SUNY Press), pp. 233–239.
- Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. (Cambridge, MA: Harvard University Press).

Editor's Note

This essay has been accepted for presentation at the XXII World Congress of Philosophy, Seoul National University, Seoul, Korea, July 30–August 5, 2008.