The Biology of Morality: Neuroscientists respond to Killen and Smetana

Biology, as we all know, is the study of life. But what is life? It turns out that scientists have yet to agree on an answer. Biology is a highly successful branch of science, and yet its central concept lacks a proper definition. Is this a scandal? That depends on your philosophical and methodological presuppositions: Which comes first, definition or understanding?

As an empiricist, I believe that we can study things like life without defining them. Of course, we must have some clue about where to start, but common-sense concepts and words suffice for that. Biologists got their start, not by rigorously defining “life,” but by studying the kinds of things that ordinary people regard as living. This strategy, I believe, works just as well for the aspect of life that we call “morality.” For empiricists, rigorously defining morality is a distant goal, not a prerequisite. If anything, I believe that defining morality at this point is more of a hindrance than a help, as it may artificially narrow the scope of inquiry.

Drs. Killen and Smetana disagree. They suggest that I and other scientists [Greene, J. et al., 2001] studying morality have gotten off on the wrong foot by failing to say what, exactly, we mean by “moral” and, worse yet, by not meaning what Kant [1785/1959] and some psychologists mean. To put it bluntly, I think that Kant’s rationalist conception of morality is mistaken and that to adopt it is to impose unnecessary limitations on one’s thinking. Rather than seeking out morality by the light of a philosopher’s definition (Kantian or otherwise), I and like-minded scientists choose to study decisions that ordinary people regard as involving moral judgment. (We have asked them for their opinions about which decisions count as “moral” decisions.) That is good enough for me. It might not be good enough for a rationalist like Kant, but I can at least take solace in the company of empiricists like Hume [1739/1978].
I personally have learned valuable lessons about the nature of morality from research in the developmental tradition [Killen and Smetana, 2006], and I hope that developmental researchers interested in morality can find value in what we, methodological empiricists, have to offer.

Joshua D. Greene
Department of Psychology
Harvard University

References

Smetana and Killen have two major concerns regarding recent affective / cognitive neuroscience regarding morality. First, that this recent literature has paid inadequate attention to the definition of morality. Secondly, that it has also deemphasized the individual’s reasons as to why an action should be considered good or bad, and focused simply instead on the individual’s judgments of whether the action is good or bad. Their points are well taken. Indeed, sometimes when the moral neuroscience literature has attempted a definition of morality it has contradicted the considerable empirical literature generated by the social domain theorists. Thus, for example, Moll and colleagues specifically argued that morality was “the sets of customs and values that are embraced by a cultural group to guide social conduct” [Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005, p. 799]; i.e., these authors effectively denied the moral and conventional distinction so well documented by the domain theorists.

The goal of this brief commentary is to consider, in the context of these concerns, how the recent moral neuroscience literature might be considered to complement and extend, rather than contradict or ignore, the social domain theorists’ earlier work. Thus, Smetana and Killen raise concern over the use of “trolley-car dilemmas” [e.g., Greene, Sommerville, Nystrom, Darley, & Cohen, 2001]; in particular, the definitional criteria used in these studies. They have stressed that the distinction made by Greene between personal and impersonal dilemmas has no counterpart in
the domain theorists’ work. While this is true, this distinction between personal and impersonal, particularly if it is re-conceptualized as between high individual victim salience and low individual victim salience [cf. Blair, Marsh, Finger, Blair, & Luo, 2006], is compatible with the domain theorist positions. According to the domain theorists, morality “includes fairness, justice, rights, and others’ welfare (e.g., when a victim is involved)” [Killen, 2007, p. 32]. One of the interesting features of the work on the trolley car problems is that under certain circumstances (the personal dilemmas), the solo victim’s welfare becomes highly salient and he/she is saved at the expense of the five other individuals. Under other circumstances (the impersonal dilemmas), this does not occur and the solo victim is sacrificed to save the five. In short, from a domain theorist position, “personal dilemma” scenarios appear to rely more on the moral domain than “impersonal dilemma” scenarios.

The moral neuroscience literature also extends the domain theorist perspective when considering the nature of morality. To consider two examples: First, much of the literature on moral development, including the domain theorists, has relatively neglected disgust-based moral transgressions. However, disgust is a powerful emotive force for attitudes towards a variety of transgressions, particularly those concerning sexual activity [Haidt, 2001]. People who hold disgust-based attitudes over actions that they consider transgressions appear to regard these transgressions as moral [Haidt & Graham, in press] and it seems likely that such people will show similar judgments regarding disgust-based transgressions as care-based transgressions (even if their justifications for the two types of transgression differ). Moreover, induced hypnotic disgust makes moral judgments more severe [Wheatley & Haidt, 2005]. Secondly, it may turn out that justice and care-based reasoning should be distinguished; that they may rely on different computational processes and should therefore be considered as two distinct social domains. The data currently is preliminary [Robertson et al., 2007] and could be at least partially explained in terms of the care-based items in that study being stronger stimuli for the moral domain. However, it is likely that the moral neuroscience literature will bring additional data to bear on this issue.

In short, one of the major contributions of the domain theorists was to consider that there were different forms of social norms: care and justice-based morality and social disorder-based convention. This was a crucial advance on the work of Piaget and Kohlberg who did not. Indeed, Piaget used data on the child’s learning of the rules of marbles to understand the development of morality and Kohlberg considered that morality developed out of more conventional reasoning. In some respects, the recent moral neuroscience literature can be considered an extension of this tradition. On the basis of this literature, we may end up distinguishing justice from care-based moral reasoning as well as coming to seriously consider disgust-based “moral” reasoning.

With respect to the second concern, that the moral neuroscience literature has deemphasized the individual’s reasoning in favor of understanding how simple judgments of whether the action is good or bad are made, there does appear to be a greater divide between the domain theorists and much of the moral neuroscience literature. According to the domain theorists, the child generates
“understandings of the social world by forming intuitive theories regarding experienced social events” [Turiel, E., Killen, M., & Helwig, 1987, p. 170]. The differing social consequences of moral and conventional transgressions result in the child’s development of separate theoretical structures (domains) for moral and conventional transgressions. In short, the domain theorists regard an individual’s verbal reasoning as critical data in understanding their moral reasoning, as it allows access to the theoretical structures upon which moral reasoning is thought to rely.

In contrast, the majority of the recent moral neuroscience positions have denied the importance of verbalizable theories and abstract reasoning more generally, arguing that judgments of permissibility are reliant on emotional responses. This claim was first made in Blair’s alternative account of the development of the moral/ conventional distinction, the Violence Inhibition Mechanism model [Blair, 1995], and particularly articulated by Haidt [2001]. Indeed, the importance of theoretical structures in basic moral reasoning does seem to be under some threat by the moral neuroscience literature; it is unclear whether an adequate account of permissibility judgments requires reference to such theoretical structures. But in an important criticism of Blair’s [1995] model, a criticism applicable to most other emotion-based positions on morality, Nichols [2004] showed that while emotional systems might generate permissibility judgments, they did not allow an individual to make immorality judgments. The same aversive affect should be present whether a victim hurt themselves or was hurt by another; i.e., this affect can generate judgments of “badness” but not “immorality” [Nichols, 2004]. It is possible that judgments of immorality, rather than simple permissibility, may require access to verbalizable theories, knowledge that a particular action is considered within the society to be wrong. In short, a failure by the moral neuroscience literature to incorporate the findings and thinking of the domain theorists is only going to result in incomplete theories.

**James Blair**  
Mood and Anxiety Program,  
National Institute of Mental Health, National Institutes of Health,  
Department of Health and Human Services, Bethesda, Maryland, USA

**References**


Haidt, J., & Graham, J. [in press]. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*.


