

Moral Reasoning on the Ground

Author(s): Richmond Campbell and Victor Kumar

Source: *Ethics*, Vol. 122, No. 2 (January 2012), pp. 273-312

Published by: The University of Chicago Press

Stable URL: <https://www.jstor.org/stable/10.1086/663980>

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Ethics*

JSTOR

# Moral Reasoning on the Ground\*

*Richmond Campbell and Victor Kumar*

---

We present a unified empirical and philosophical account of moral consistency reasoning, a distinctive form of moral reasoning that exposes inconsistencies among moral judgments about concrete cases. Judgments opposed in belief or in emotion and motivation are inconsistent when the cases are similar in morally relevant respects. Moral consistency reasoning, we argue, regularly shapes moral thought and feeling by coordinating two systems described in dual process models of moral cognition. Our empirical explanation of moral change fills a gap in the empirical literature, making psychologically plausible a defensible new model of justified moral change and a hybrid theory of moral judgment.

---

After Germany occupied Norway in World War II, Jan Baalsrud, a young Norwegian resistance fighter who was pursued by the Nazis, stumbled into a home in a small village.<sup>1</sup> Having barely survived an avalanche, he was snow blind and near death from cold and exhaustion. The family there faced the choice of turning him over to the Nazis or trying to get him well enough to attempt an escape, with the strong possibility that their efforts would be discovered and the whole village put to death in retaliation. Jan formed a friendship with Marius, a young man in the family. Marius favored helping Jan escape. Although Marius's mother pitied Jan and wanted to save him as a

\* Early versions of this article were presented at Dalhousie University, the University of Arizona, Harvard University, and the 2011 American Philosophical Association Pacific meeting in San Diego. For constructive criticism and suggestions, we are grateful to seven anonymous referees and editors of this journal and John Barresi, Kristin Borgerson, David Braybrooke, Fiery Cushman, Dan Demetriou, Brian Fiala, Jeremy David Fix, Joshua Greene, Ami Harbin, Terry Horgan, Michael Hymers, Chike Jeffers, Ben Kearns, Andrew Kernohan, Colin Klein, Levi Lauer, Theresa Lopez, Duncan MacIntosh, Shaun Nichols, Eve Roberts, Sue Sherwin, Zeynep Soysal, Frans Svensson, Mark Timmons, and Meghan Winsby. Victor Kumar's work on the essay was supported by a fellowship from the Social Sciences and Humanities Research Council of Canada.

1. The story, based on firsthand accounts, is taken from David Howarth, *We Die Alone* (New York: Macmillan, 1955), 98–114.

*Ethics* 122 (January 2012): 273–312

© 2012 by The University of Chicago. All rights reserved. 0014-1704/2012/12202-0004\$10.00

fellow patriot, her strong sense of responsibility for her family and her community made her at first opposed to her son's plan. It would be moral madness, she thought, to stake everyone else's life on the small chance of saving Jan's. Then Marius put to his mother the following question: "What ought a family in Oslo to do if I were in Jan's shoes? Should they give me to the Nazis?"

Seeing their situation in light of this example helped her think of Jan as a boy like her son rather than as a stranger, and in the end she consented to her son's wishes. Is there a logic to her change in feelings? Marius's mother has but four choices. She could bite the bullet and say, out of consistency, that Marius should be released to the Nazis, but she rejects this option. We are not told why but will suppose that she finds morally repugnant the prospect of her son being used in this way to appease the Nazis. Having rejected this option, however, she can find no morally relevant difference in the situations that would save her from inconsistency. She cannot say, for example, that Jan is not her son, for the hypothetical family in Oslo could make the symmetrical claim about Marius. But, importantly, it is also not a viable response to Marius's challenge to dismiss the apparent inconsistency. Not to care about being consistent in one's moral attitudes and feelings in this life-and-death choice would undermine one's credibility as a moral agent, not to mention as a trustworthy and responsible parent; one's moral responses would be unpredictable and one's character unreliable. The fourth option is, of course, to change her mind, and her family, who well understood the force of Marius's challenge since it could apply as easily to them, supported this decision. Fortunately they managed to nurse Jan back to health and help him escape under the noses of the occupation forces, although the weight of their decision haunted Marius's mother for the rest of her life.

Marius's implicit argument exemplifies a norm of consistency—treating like cases alike. "Moral consistency reasoning," as we call it, exposes latent moral inconsistencies, embodied in conflicting moral judgments about cases that are, by one's own lights, similar in morally relevant respects. Consistency reasoning is frequently employed in applied ethics and the law. But, as the example above and others to follow illustrate, it is also common in everyday moral thought and dialogue. Consistency reasoning influences not just our moral opinions but also our intuitive and emotional moral dispositions. It changes what we think, but it also changes how we feel. Dialogical appeals to consistency have likely been a source of moral progress. For example, the socially enforced demand to apply norms of fairness and justice consistently has helped engender progressive intuitions about the moral status of women, persons of color, and other groups

that historically have been regarded as unworthy of respect. On the one hand, there is an empirical question about the cognitive and social mechanisms through which consistency reasoning has brought about moral change. On the other hand, there is a philosophical question about the value of consistency reasoning and its role in moral justification. As naturalists, we think it fruitful to pursue answers to each question in light of the other.

It is not immediately obvious how consistency reasoning can be accommodated within existing empirical theories of moral cognition. A popular class of theories that seeks to capture intuitive, emotional moral processing depicts moral reasoning as largely post hoc and epiphenomenal. According to these theories, moral evaluation is guided by a specialized, affective, intuitive system and, much less frequently, by a general-purpose, deliberative, reasoning system. Proponents have so far ignored whether reasoning may over time influence the operation of our affective, intuitive system. They admit, furthermore, that “we know little about how the intuitive/affective and conscious/cognitive systems interact on-line in the production of moral judgments.”<sup>2</sup> We suggest that the affective, intuitive system and the deliberative, reasoning system work together in moral consistency reasoning to elicit long-term emotional moral change.

We begin the essay with the empirical aim of understanding moral consistency reasoning in the context of a dual process model of moral cognition that does not entail any of the popular theories just sketched. Eventually, however, we turn to the normative import of our empirical theorizing. An empirical account of moral consistency reasoning explains social moral change and, we argue, enriches philosophical models of moral justification. Furthermore, an empirical account of consistency reasoning and the functional role it plays in our social lives can help explain why it is worth engaging in.

In Section I we describe a minimalist dual process model of moral cognition and the need to examine diachronic change within our affective, intuitive system. In Section II we describe moral consistency reasoning through examples and experimental research, suggesting that it is a mechanism of reason-based moral change. In Section III we present a new integrated dual process model of moral consistency reasoning. Unlike other forms of moral reasoning that do not engage intuition, moral consistency reasoning emerges from coordinated activity across the intuitive system and the reasoning system. In Section IV we characterize the functional role of moral consistency reasoning

2. Fiery Cushman, Liane Young, and Joshua Greene, “Multi-System Moral Psychology,” in *The Moral Psychology Handbook*, ed. J. Doris (Oxford: Oxford University Press, 2010), 47–71, 67.

as a deliberative, typically social activity that secures the cooperative benefits of morality by exposing and eliminating practical inconsistencies among our moral emotional responses.

Finally, in Section V we explore the relevance of our empirical theorizing for normative ethics and metaethics. We argue that our empirical theory of moral consistency reasoning explains the psychological feasibility of a defensible new normative model of moral justification. We argue too that our more comprehensive understanding of moral reasoning fits best with a hybrid cognitivist-expressivist theory of moral judgment.

## I. DUAL PROCESS MODEL AND MORAL CHANGE

Dual process models have found traction in many areas of cognitive and social psychology.<sup>3</sup> Jonathan Haidt, Joshua Greene, and other psychologists have recently advanced controversial theories of moral cognition, all of which have in common but reach far beyond a minimalist moral dual process model (MMDP).<sup>4</sup> MMDP states that two distinct systems guide moral evaluation. Each system is associated with a set of characteristic features. System 1, a domain-specific capacity, is fast, automatic, effortless, affective, unconscious, and impenetrable. System 2, our domain-general capacity for reasoning, is slow, controlled, effortful, “cognitive,” conscious, and penetrable (see table 1).<sup>5</sup> Some of these features are self-explanatory; others require a brief comment. Moral system 1 is one of perhaps many domain-specific capacities, affective in the sense that it involves emotional processing; system 2

3. Keith Stanovich and Richard West, “Individual Differences in Reasoning: Implications for the Rationality Debate?” *Behavioral and Brain Sciences* 23 (2000): 645–65; Shelly Chaiken and Yaacov Trope, eds., *Dual Process Theories in Social Psychology* (New York: Guilford, 1999).

4. Jonathan Haidt, “The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment,” *Psychological Review* 108 (2001): 814–34, and “The New Synthesis in Moral Psychology,” *Science* 316 (2007): 998–1002; Joshua Greene, “The Secret Joke of Kant’s Soul,” in *Moral Psychology*, vol. 3, ed. W. Sinnott-Armstrong (Cambridge, MA: MIT Press, 2008), 35–79; Joshua Greene and Jonathan Haidt, “How (and Where) Does Moral Judgment Work?” *Trends in Cognitive Science* 6 (2002): 517–23; Joshua Greene, Sylvia Morelli, Kelly Lowenberg, Leigh Nystrom, and Jonathan Cohen, “Cognitive Load Selectively Interferes with Utilitarian Moral Judgment,” *Cognition* 107 (2008): 1144–54; Cushman, Young, and Greene, “Multi-System Moral Psychology.” For a recent critical review of this literature and its implications for normative moral theory, see Selim Berker, “The Normative Insignificance of Neuroscience,” *Philosophy and Public Affairs* 37 (2009): 293–329; and Victor Kumar and Richmond Campbell, “On the Normative Significance of Experimental Moral Psychology,” *Philosophical Psychology*, forthcoming.

5. A dual process model of some cognitive capacity does not entail that our mind consists in just these two kinds of systems. Some mental processes may resist neat categorization into processes characteristically produced by system 1 or system 2.

TABLE 1

LIST OF FEATURES CHARACTERISTIC OF EACH SYSTEM  
IN A MINIMALIST MORAL DUAL PROCESS MODEL

The Intuitive System (System 1)	The Reasoning System (System 2)
Domain specific	Domain general
Affective	Cognitive
Fast	Slow
Automatic	Controlled
Effortless	Effortful
Impenetrable	Penetrable
Unconscious	Conscious

is “cognitive” only in the sense that the processing does not involve emotions or other motivational states. Information processing takes place in both systems. In system 1 it is unconscious, while some but not all of the processing in system 2 is conscious. System 1 is “impenetrable” (or “encapsulated”) in a certain technical sense; that is, cognitive processes do not influence the online operation of system 1. If an action seems intuitively wrong, it will still seem that way even if one overrides one’s intuition with a cognitive judgment that the action is morally permissible.

We want to be clear from the start that although we endorse a dual process model of moral cognition, our understanding of the model is minimalist, including only the features summarized in table 1. For example, some dual process theorists hold that system 1 is associative, system 2 rule governed; system 1 is innate, system 2 learned; and so on.<sup>6</sup> MMDP is neutral with respect to further such claims. Note, in particular, Greene’s view that deontology emerges from system 1, while utilitarianism emerges from system 2, and that system 1 is epistemically suspect, while system 2 is “rational.”<sup>7</sup> As we understand it, MMDP has no such implications. Intuitive moral judgments are not simply blunt emotional reactions but are guided by moral norms, even when subjects do not have conscious access to them. Although we share with Haidt and Greene a commitment to MMDP, there is not much else in common between our view of moral cognition and theirs. Specific points of disagreement will emerge as we proceed.

6. A less minimalist list of various properties in each cluster is provided in Haidt, “Emotional Dog and Its Rational Tail,” 818.

7. Greene, “Secret Joke of Kant’s Soul.” For criticism, see Kumar and Campbell, “On the Normative Significance of Experimental Moral Psychology.”

Drawing on the research of Haidt and Greene as well as others, we begin by summarizing evidence for MMDP, according to which moral evaluation is guided by both an affective, intuitive system specific to the moral domain and a general purpose reasoning system. Dual process theorists tend not to pay much attention to moral change. However, we believe that intuitive, emotional processing is intelligent and flexible, a position that can be articulated more precisely in terms used above to characterize the difference between system 1 and system 2.<sup>8</sup> System 1 is *impenetrable* in the sense that controlled cognition has no immediate effect on its internal operation and the outputs it yields. However, it does not follow that cognition has no long-term influence on system 1. In short, while system 1 is as a matter of definition *synchronically* impenetrable, it may nevertheless be *diachronically* penetrable. Our aims in this section are, first, to rehearse the evidence in support of MMDP and, second, to suggest that moral system 1 is diachronically penetrable. In later sections we focus on a mechanism that underlies such change.

Historically, since at least Piaget, research on moral cognition has focused on moral reasoning.<sup>9</sup> More recent psychological work targets intuitive moral appraisal. Jonathan Baron cites empirical findings suggesting that some moral judgments violate norms and values that subjects explicitly accept.<sup>10</sup> In Haidt and colleagues' work on "moral dumbfounding," subjects persist in holding moral opinions not supported by the principles they subsequently offer in support of those opinions.<sup>11</sup> The judgments appear to be produced quickly and by processes to which the subjects lack conscious access. Work by researchers at Harvard is consistent with these findings. A study conducted by Fiery

8. A similar view of moral intuition is defended by James Woodward and John Allman, "Moral Intuition: Its Neural Substrates and Normative Significance," *Journal of Physiology—Paris* 101 (2007): 179–202; John Allman and James Woodward, "What Are Moral Intuitions and Why Should We Care about Them? A Neurobiological Perspective," *Philosophical Issues* 18 (2008): 164–85.

9. Lawrence Kohlberg's influential theory of moral development presents humans as engaging in different kinds of moral reasoning at successive stages of moral development, reaching conclusions about right and wrong first by appeal to authority and convention and later through universal principles of justice. Kohlberg, "The Development of Children's Orientation toward a Moral Order," pt. 1, "Sequence in the Development of Moral Thought," *Vita Humana* 6 (1963): 11–33; Elliot Turiel, *The Development of Social Knowledge: Morality and Convention* (Cambridge: Cambridge University Press, 1983); cf. Carol Gilligan, *In a Different Voice: Psychological Theory and Women's Development* (Cambridge, MA: Harvard University Press, 1982).

10. Jonathan Baron, "Nonconsequentialist Decisions," *Behavioral and Brain Sciences* 17 (1994): 1–42.

11. Jonathan Haidt, Frederik Bjorklund, and Scott Murphy, "Moral Dumbfounding: When Intuition Finds No Reason" (unpublished manuscript, University of Virginia, 2000). We criticize this view in the next section.

Cushman and colleagues indicates that subjects are guided by something like the principle of double effect—roughly, the principle that says harms are worse if intended than if foreseen but unintended—even though they cannot articulate that principle.<sup>12</sup> All of this suggests moral cognition is as much a matter of unconscious intuition as it is reasoning. As the work by Cushman et al. suggests, the intuitive system, although unconscious in its operation, is guided by norms, some quite sophisticated.

Greene has gathered experimental evidence for a dual process model at odds with the model we endorse.<sup>13</sup> “Deontological” judgments, he argues, tend to be produced by system 1, whereas “consequentialist” judgments tend to be produced by system 2—where these terms appropriated from ethical theory are defined operationally. As Greene uses the terms, “deontological” judgments are those substantive conclusions associated with deontological theories, for example, that lying is wrong even if it produces a better outcome; “consequentialist” judgments are those substantive conclusions associated with consequentialist theories, for example, that it is right to sacrifice one person to save many others. By assaying subjects “deontological” and “consequentialist” judgments, Greene and colleagues attempt to probe the inner workings of systems 1 and 2. We deny, however, that there is a general pattern linking system 1 with “deontological” judgments and system 2 with “consequentialist” judgments. After rehearsing the relevant studies, we will argue that they support only MMDP and not Greene’s more expansive theory of moral cognition.

On Greene’s view, we use system 1 to respond quickly and automatically to intentional harm and rights violations; it is as if these conditions trigger an internal alarm. By contrast, when we are concerned with what will produce the most utility, system 2 is enlisted to make the relevant comparisons and calculations. Greene argues that reaction time data suggest this pattern and that further evidence is found in studies in which a cognitive load is imposed on subjects.<sup>14</sup>

12. Fiery Cushman, Liane Young, and Marc Hauser, “The Role of Conscious Reasoning and Intuition in Moral Judgment,” *Psychological Science* 17 (2006): 1082–89; Marc Hauser, Fiery Cushman, Liane Young, R. Kang-Xing Jin, and John Mikhail, “A Dissociation between Moral Judgments and Justifications,” *Mind and Language* 22 (2007): 1–21.

13. Greene, “Secret Joke of Kant’s Soul.”

14. Compare Jonathan McGuire, Robyn Langdon, Max Coltheart, and Catriona Mackenzie, “A Reanalysis of the Personal/Impersonal Distinction in Moral Psychology Research,” *Journal of Experimental Social Psychology* 45 (2009): 577–80; Berker, “Normative Insignificance of Neuroscience,” 308–11. For responses to these criticisms, see Joshua Greene, “Dual Process Morality and the Personal/Impersonal Distinction: A Reply to McGuire, Langdon, Coltheart and Mackenzie,” *Journal of Experimental Social Psychology* 45 (2009): 581–84, and “Notes on ‘The Normative Insignificance of Neuroscience’ by Selim Berker,” <http://www.wjh.harvard.edu/~jgreene/GreeneWJH/Greene-Notes-on-Berker-Nov10.pdf>.



Quite generally, system 2, but not system 1, tends to be impeded when subjects are asked to concurrently perform other tasks. Greene et al. find that a cognitive load slows down “consequentialist” judgments but not “deontological” judgments, inferring that the former are generated by system 2, the latter by system 1.<sup>15</sup>

Neuroscientific research that takes advantage of this pattern appears to yield further support for Greene’s dual process model, suggesting that moral judgment is realized in two different areas of the brain. In an fMRI study by Greene et al., “consequentialist” judgments requiring calculation are correlated with increased activation in areas of the brain associated with working memory; “deontological” judgments are correlated with areas of the brain associated with social/emotional cognition.<sup>16</sup> A further piece of neuroscientific evidence emerges from work by Koenigs et al. on patients with damage to the ventromedial prefrontal cortex (VMPC), an area of the brain associated with social cognition and emotion.<sup>17</sup> VMPC patients tend to exhibit an abnormally consequentialist pattern of judgment, suggesting that system 1 has been damaged, while system 2 remains intact. These two studies point to a neurophysiological dissociation between reasoning and intuition, thus seeming to provide further support for Greene’s dual process model.

Contrary to Greene, the distinction between system 1 and system 2 does not line up with the distinction between “deontological” judgments and “consequentialist” judgments. First, many characteristically consequentialist judgments are emotion based and produced by system 1, for example, judgments that an action that produces some massive, unmitigated amount of misery is wrong. Indeed, Greene’s own neuroimaging studies reveal that an area of the brain associated with emotion, the posterior cingulate, is activated when subjects make

15. Greene et al., “Cognitive Load Selectively Interferes with Utilitarian Moral Judgment.”

16. Joshua Greene, R. Brian Sommerville, Leigh Nystrom, John Darley, and Jonathan Cohen, “An fMRI Investigation of Emotional Engagement in Moral Judgment,” *Science* 293 (2001): 2105–8; Joshua Greene, Leigh Nystrom, Andrew Engell, John Darley, and Jonathan Cohen, “The Neural Bases of Cognitive Conflict and Control in Moral Judgment,” *Neuron* 44 (2004): 389–400.

17. Michael Koenigs, Liane Young, Ralph Adolphs, Daniel Tranel, Fiery Cushman, Marc Hauser, and Antonio Damasio, “Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments,” *Nature* 446 (2007): 908–11. For critical discussion, see Jorge Moll and Ricardo de Oliveira-Souza, “Moral Judgments, Emotions and the Utilitarian Brain,” *Trends in Cognitive Science* 11 (2007): 319–21; Joshua Greene, “Why Are VMPFC Patients More Utilitarian? A Dual-Process Theory of Moral Judgment Explains,” *Trends in Cognitive Science* 11 (2007): 322–23; Jorge Moll and Ricardo de Oliveira-Souza, “Response to Greene: Moral Sentiments and Reason; Friends or Foes?” *Trends in Cognitive Science* 11 (2007): 323–24.

consequentialist judgments, thus implicating system 1.<sup>18</sup> System 2 seems to be at work in consequentialist judgments only when subjects must compare or calculate the relative weight of different values, and it is these kinds of cases that Greene exploits in his experimental work. Even here, as just noted, system 1 is engaged since emotions are at play. A second reason the pattern is disrupted is that deontological judgments are often not the product of system 1's "internal alarms." When presented with difficult moral problems, individuals routinely arrive at a characteristically deontological conclusion after explicitly applying deontological principles in system 2. The pattern that Greene exploits is thus only local—it obtains only in the moral dilemmas he presents to his subjects. A recent study by Kahane et al. confirms this interpretation.<sup>19</sup> Employing a wider range of moral dilemmas, the authors find that the neurological correlates of system 1 are associated with "intuitive" judgments, while the neurological correlates of system 2 are associated with "counterintuitive" judgments.

We think that Greene's experimental evidence supports MMDP (as defined above). Suppose MMDP is true. Then moral dilemmas and other hard cases—including those that involve cost-benefit trade-offs—would engage system 2 as well as system 1. Hence, they would be slower in general and under cognitive load. Moreover, other cases that do not require careful reasoning—in Greene's dilemmas they elicit "deontological" judgments—would be made by system 1 alone and infused with affect, as the neuroscientific evidence indicates. Given that dual process theorizing has been successful in other areas of psychology and given that MMDP explains these results, it is a plausible model, especially once it is clearly differentiated from Greene's controversial elaboration of dual process. While MMDP is plausible, much more empirical work needs to be done, as we will indicate; an alternative model may prove to be superior in the end.<sup>20</sup> Our commitment to MMDP, although well supported, must be tentative.

Having rehearsed the evidence in support of MMDP, we now turn

18. Greene et al., "Neural Bases of Cognitive Conflict and Control in Moral Judgment."

19. Guy Kahane, Katja Wiech, Nicholas Shackel, Miguel Farias, Julian Savulescu, and Irene Tracey, "The Neural Basis of Intuitive and Counterintuitive Judgments," *Social Cognitive and Affective Neuroscience*, electronically published March 18, 2011, doi:10.1093/scan/nsr005.

20. For example, a diachronic structure is also implied by a skill-based analysis of moral competence that does not entail dual process. While we believe that such an analysis is appropriate for understanding changes in the speed and accuracy at which information is processed in system 2, it cannot account for the system 1 character of moral responses that appear in very early stages of moral development. We cannot pursue this matter here.

to an issue that dual process theorists have so far failed to address. To explain precisely what we have in mind, we begin by discussing a locus classicus for dual process theory, Jerry Fodor's 1983 book *Modularity of Mind*.<sup>21</sup> Fodor argues that perceptual systems and central cognitive systems can be distinguished roughly along the lines that have come to distinguish system 1 and system 2. Of central interest to us is Fodor's main argument that perceptual systems are impenetrable. Fodor appeals to visual illusions like the Muller-Lyer figure, in which two parallel lines appear to be of different lengths; even after we realize that the lines are the same length, it continues to appear visually that one is longer than the other. Our explicit beliefs, Fodor infers, cannot penetrate the processing in our visual system. Similarly, according to MMDP, explicit moral beliefs cannot penetrate moral system 1. For example, although utilitarians may reason that one is morally obliged to shoot one of the captured natives in Bernard Williams's famous thought experiment, it can still feel intuitively wrong to them to shoot one of the natives.<sup>22</sup>

Fodor's critics have argued that the processing of visual illusions is less uniform than he suggests.<sup>23</sup> Cross-cultural psychological research begun by W. H. R. Rivers and replicated since indicates that susceptibility to visual illusions is not static.<sup>24</sup> It turns out that whether individuals see the two lines in the Muller-Lyer figure as differing in length depends on whether they have lived in a "carpentered" environment—that is, in environments that contain walls, buildings, furniture, and other objects with sharp right angles. In the visual systems of those who have lived in uncarpentered environments, the arrows on the end of the lines in the Muller-Lyer figure are not processed as right angles and thus do not serve as depth cues. Thus, it seems that the visual system is penetrable—not synchronically but diachronically.

Might moral system 1 be likewise diachronically (although not synchronically) penetrable? Citing empirical studies of normative but nonmoral intuitions, John Allman and James Woodward suggest that

21. Jerry Fodor, *The Modularity of Mind* (Cambridge, MA: MIT Press, 1983).

22. Bernard Williams, "A Critique of Utilitarianism," in *Utilitarianism: For and Against*, ed. J. J. C. Smart and B. Williams (Cambridge: Cambridge University Press, 1973), 77–150, 98–100.

23. Robert McCauley and Joseph Henrich, "Susceptibility to the Muller-Lyer Illusion, Theory-Neutral Observation and the Diachronic Penetrability of the Visual Input System," *Philosophical Psychology* 19 (2006): 1–23.

24. W. H. R. Rivers, "Observations on the Senses of the Todas," *British Journal of Psychology* 1 (1905): 321–96; Marshall Segall, Donald Campbell, and Melville Herskovits, "Cultural Differences in the Perception of Geometric Illusions," *Science* 139 (1963): 769–71.

“emotional processing and the structures underlying moral intuition can be heavily influenced by learning and experience.”<sup>25</sup> For example, experienced nurses rely on their feelings in hard cases to make judgments about whether a patient is in trouble that are more accurate than those of their more novice colleagues; subjects given the Iowa Gambling Task eventually are attracted to the decks of cards that will yield a higher net total of rewards over time, while patients with normal memory and calculating ability who have damage to the VMPC (associated with social cognition and emotion) are not and consistently make poor choices; other studies relate similar effects.<sup>26</sup>

The research Allman and Woodward cite provides only indirect evidence that the moral system 1 is diachronically penetrable. Further experimental research on moral intuition would be valuable; most existing research on moral change has been conducted within the reasoning-centered approaches developed by Lawrence Kohlberg and his students.<sup>27</sup> Nevertheless, it is highly plausible that moral intuitions change over time, both over the course of history and in individual development. In a recent issue of *Nature*, Paul Bloom criticizes existing dual process approaches precisely for failing to account for change in not just moral belief but also moral intuition: “Emotional responses alone cannot explain one of the most interesting aspects of human nature: that morals evolve. The extent of the average person’s sympathies has grown substantially and continues to do so. Contemporary readers of *Nature*, for example, have different beliefs about the rights of women, racial minorities and homosexuals compared with readers in the late 1800s, and different intuitions about the morality of practices such as slavery, child labour and the abuse of animals for public entertainment. Rational deliberation and debate have played a large part in this development.”<sup>28</sup>

In the next few sections, we offer an explanation of reason-based moral change. We applaud Bloom’s call for research on this subject, but we also believe that he implicitly takes on board a false dichotomy between reason and emotion. Dual process theorists like Haidt and Greene generally suppose that moral judgment is either emotion-driven intuition or coolheaded reasoning; reasoning, because it is coolheaded, cannot exert much control over our affectively charged moral commitments. Moral consistency reasoning is an engine of

25. Allman and Woodward, “What Are Moral Intuitions and Why Should We Care about Them?” 169.

26. *Ibid.*, 170–72.

27. Elliot Turiel, *The Culture of Morality: Social Development, Context, and Conflict* (Cambridge: Cambridge University Press, 2002).

28. Paul Bloom, “How Do Morals Change?” *Nature* 464 (2010): 490.

moral change; however, its power lies precisely in its ability to recruit reason and emotion in the cooperative interaction of systems 1 and 2.

## II. MORAL CONSISTENCY REASONING AND MORAL CHANGE

In the next section we will present a dual process model of how reason and emotion interact in moral consistency reasoning, but at this point we need a better grasp of how consistency reasoning is exemplified in examples. Consider again the example with which we opened the article. Marius's mother judges initially that she and her family are morally obliged to turn over Baalsrud, the young freedom fighter whom the Nazis are pursuing. As much as she sympathizes with Baalsrud and his cause, the risks to her family and her community are too great. Marius shows her, however, that her initial response is inconsistent with her other moral values. Knowing his mother's feelings, Marius asks what she thinks ought to be done if he found himself in the same situation in another community. Unable to find a morally relevant difference between the two situations, she comes to the conclusion that, in fact, she and her family ought to help Baalsrud.

In this example, incompatible moral responses toward two separate instances of the same type of situation are brought into clear view. Note that the moral responses are in part emotional and motivational. Their obvious incompatibility as responses to the same type of situation would be normally undesirable, for reasons we will elaborate in Section IV. In successful consistency reasoning, one response is independently less tenable or easier to relinquish than the other because of already established patterns of moral response, and unless a relevant difference between the two instances is found, a person engaged in this reasoning must either accept inconsistency and thereby give up any semblance of having a good reason for either of the responses or else revise the less tenable response. There will always be differences between any two instances, and theoretically any difference could be claimed to be relevant, but since the pressure to maintain consistency transfers to any difference cited, there may be no difference that can be consistently maintained as relevant.

G. E. Hughes illustrates this last point with a biblical example.<sup>29</sup> After learning of King David's betrayal and deceit in allowing Bathsheba's husband to be isolated and killed in battle so that David could marry Bathsheba, the prophet Nathan told David the story of a wealthy landowner who deceived and betrayed a poor shepherd who

29. G. E. Hughes, "Moral Condemnation," in *Essays in Moral Philosophy*, ed. A. I. Melden (Seattle: University of Washington Press, 1958), 108–34, 125.

was passing through his lands by inviting the shepherd to dine with him and then having the guest unwittingly eat his pet lamb. David was outraged upon hearing this story and demanded to know the identity of the man who had done this morally repugnant thing, at which point Nathan said: "Thou art the man." In this case there are a multitude of differences that could be cited, but none were seen, even by David, to be relevant. David might have claimed that he was king, not simply a wealthy landowner, but it would be easy to change the original example so that a king had committed the betrayal of trust. As in the case of Marius's mother, the option of biting the bullet, in this case saying that betraying the poor shepherd in this way was okay, was not morally viable for David and presumably not for those who had put their trust in his moral authority, as he well knew. Dismissing inconsistency as unimportant would equally undermine his moral authority. Consistency drove him to condemn his own actions and character.

Evidence that the capacity for consistency-based change in moral attitude is widespread can be found in experimental work on trolley cases. Two widely discussed versions of the trolley case have perplexed philosophers. In the bystander case, we are asked to judge whether it is permissible to flip a switch so that a trolley runs over one person instead of five; many judge that it is permissible. In the footbridge case, we are asked to judge whether it is permissible to push a fat man off a footbridge and into the path of the trolley, sacrificing this one person to save five others; many judge that so doing is impermissible. The two judgments are philosophically perplexing since the differences between the two cases may seem morally irrelevant. According to some philosophers, the principle of double effect licenses our opposing judgments, while for others that principle is not defensible, and the cases should be treated alike. Now, our concern at this point is not with the justification of moral judgments but in characterizing their etiology. However, philosophers' reasoning about trolley cases is an instance of the kind of reasoning in which we are interested. And, strikingly, it appears that subjects who are given both the bystander case and the footbridge case are disposed to spontaneously engage in such reasoning. Those given the footbridge case first tend to judge that it is wrong to push the fat man, as usual. But then when given the bystander case, they tend to say it is wrong to flip the switch.<sup>30</sup> A natural explanation for this result is that subjects perceive

30. Eric Schwitzgebel and Fiery Cushman, "Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-philosophers," *Mind and Language*, forthcoming; see also Lewis Petrinovich and Patricia O'Neill, "Influence of Wording and Framing Effects on Moral Intuitions," *Ethology and Sociobiology* 17 (1996): 145–71.

that there are no morally relevant differences between the two cases and infer that whatever one should do in the one case, one should also do in the other.<sup>31</sup>

Later we will provide a general analysis of these examples. At this point, we want to suggest that consistency reasoning, besides being at work in these examples and others like it, has played a significant role in long-term, large-scale emotional moral change, for example, in changes to sexist and racist attitudes and the affectively laden intuitions that support those attitudes. Although we see no reason to deny that moral reasoning is, more often than we think, merely rationalization of emotion-driven moral intuitions, we believe that Haidt and Greene ignore an important way in which moral reasoning guides moral judgment and changes how the intuitive system automatically processes information. Haidt and Greene have been led to regard moral reasoning as rationalization in part because the methodology of much (but not all) experimental research tends to be temporally limited.<sup>32</sup> Often, the effect of reasoning on moral judgment emerges

31. The opposite order effect does not generally obtain. Those who are given the bystander case first, and say that it is OK to flip the switch, go on to say, as usual, that it is wrong to push the fat man from the footbridge. Arguably, the judgment about the footbridge case is firmer than the judgment about the bystander case. Thus, subjects are less likely to revise their initial opinion about the footbridge case. It would be interesting to know what would happen if subjects were given the chance to reconsider the bystander case. Is the desire to be consistent so strong that they would change their minds about a case about which they had already registered a position?

32. In experiments by Haidt et al., e.g., subjects are presented with moral narratives that evoke an emotional response but in which the usual justifications for moral appraisal are unavailable; subjects arrive at a moral opinion and search without success for different ways to justify it. See Haidt, Bjorklund, and Murphy, "Moral Dumbfounding"; Cushman, Young, and Hauser, "Role of Conscious Reasoning and Intuition in Moral Judgment"; Hauser et al., "Dissociation between Moral Judgments and Justifications." According to Haidt's social intuitionist model, moral reasoning occasionally influences our moral opinions, but only when intuition is weak or when such reasoning reinforces group conformity (Haidt, "Emotional Dog and Its Rational Tail," 818–19). For empirical criticism of Haidt's model, see David Pizarro and Paul Bloom, "The Intelligence of the Moral Intuitions: Comment on Haidt," *Psychological Review* 110 (2003): 193–96; Herbert Saltstein and Tziporah Kasachkoff, "Haidt's Moral Intuitionist Theory: A Psychological and Philosophical Critique," *Review of General Psychology* 8 (2004): 273–82; Cordelia Fine, "Is the Emotional Dog Wagging Its Rational Tail, or Chasing It?" *Philosophical Explorations* 9 (2006): 83–98; Darcia Narvaez, "The Social Intuitionist Model and Some Counter-Intuitions," in *Moral Psychology*, vol. 2, *The Cognitive Science of Morality: Intuition and Diversity*, ed. W. Sinnott-Armstrong (Cambridge, MA: MIT Press, 2008), 233–40. Greene's view of moral reasoning, as we have seen, is somewhat different. For him, reasoning in support of "deontological" judgments is post hoc rationalization, but reasoning can guide judgments when they are confronted with dilemmas that involve competing values and bring them to engage in cost-benefit analysis. Despite these differences, neither Haidt nor Greene examines the effect of moral reasoning on emotional moral responses over any significant length of time.

only over a significant length of time and not, for example, in one-shot consideration of moral dilemmas. Our emotion-driven intuitive system may well synchronically fix moral judgment and drive moral reasoning. But consistency reasoning has a diachronic effect on moral judgment and moral intuition.

Consider, by way of illustration, the changes many straight persons undergo in their moral attitudes toward homosexuality. Influenced by their parents, their peers, and perhaps also the apparent religious authority of those perspectives, many begin adolescence thinking that gay sex and sexuality are disgusting and wrong. However, in time they may come to doubt the authority of religious condemnation of homosexuality, either because they no longer have faith or because they interpret their faith differently. They may also become aware of the immense variety of sexuality in nature, no longer associating gay sex with what is unnatural. Still, feelings of moral disgust instilled early on may remain. Then they reflect. What possible difference between straight and gay sex could justify the difference in moral feeling? What is valuable about many heterosexual relationships is the mutual love, care, and respect they embody, the pleasure derived from them; obviously none of this is exclusively heterosexual. Homosexual love is not procreative, but neither is most heterosexual love, and certainly the latter is permissible in the absence of religious dogma. Further, some kinds of sexual activities are thought to be perverse and wrong because they cause harm to children, or are exploitative in some way, but homosexual activity is not like these. In sum, the demand to make our moral responses to different cases (to gay sex, to straight sex, to sexual perversion) consistent with one another puts pressure on negative moral attitudes to homosexuality. As a result of reflection over a significant duration, many persons who were raised to condemn gay sex come to believe that homosexuality is morally okay. They eventually lose, to a significant degree, the emotional response associated with their old, naive moral perspective, no longer feeling intense disgust when they see or think about homosexual activity. Even if the change is not complete,<sup>33</sup> their moral feelings toward those with a different sexual orientation have changed, in part because of the repeated discomfort they experience whenever they are confronted with inconsistency in their moral feelings about sexual difference.

What is the significance of consistency-based moral emotional

33. Work on implicit racial attitudes suggests that many prejudices are deeply entrenched and cannot be overcome completely. See, e.g., Patricia Devine, "Stereotypes and Prejudice: Their Automatic and Controlled Components," *Journal of Personality and Social Psychology* 56 (1989): 5–18.



change in the context of the current literature? First, Haidt and Greene accord moral reasoning little power to influence moral judgment when intuitive reactions are strong, whereas the examples above indicate that consistency reasoning is regularly operative in moral judgment.<sup>34</sup> Second, since Haidt and Greene hold in effect that reason is slave to emotion, they are silent about the possibility of reasoning influencing the operation of the intuitive system. In particular, they are silent about diachronic emotional moral change based on consistency reasoning. According to Haidt's Social Intuitionist Model, social reasoning can affect moral judgment via motives for group conformity, but consistency reasoning is quite unlike this arational process.<sup>35</sup> Strikingly, the motive to be consistent can keep one from conforming to the group, as when one bravely calls attention to and seeks to resolve an inconsistency that others in one's community refuse to face.

Consistency arguments abound in everyday moral discussion but also feature in moral arguments by philosophers. Peter Singer's well-known argument for making charitable contributions to reduce famine as well as his case for becoming vegetarian turn critically on consistency reasoning that has fundamentally the same structure.<sup>36</sup> In these cases, consistency reasoning draws on utilitarian intuitions (e.g., that one should save the child from drowning in a shallow pond at the cost of getting one's suit dirty), but other philosophers draw on deontological intuitions, as in the famous defense of abortion by Judith Jarvis Thomson.<sup>37</sup> Thomson pumps our intuitions about abortion and the alleged primacy of the right to life by having us think about the conditions under which it is permissible to unplug a famous violinist from one's body. The force of the appeal to consistency in these

34. See Pizarro and Bloom, "Intelligence of the Moral Intuitions"; and the reply in Jonathan Haidt, "The Emotional Dog Does Learn New Tricks: A Reply to Pizarro and Bloom," *Psychological Review* 110 (2003): 197–98. Pizarro and Bloom argue that reasoning can exert control over moral intuitions through (1) cognitive appraisals that change the significance of descriptive facts and (2) control over the input on which our intuitions operate. Moral consistency reasoning is a different phenomenon.

35. Haidt, "Emotional Dog and Its Rational Tail," 817–19.

36. Peter Singer, "Famine, Affluence, and Morality," *Philosophy and Public Affairs* 1 (1972): 229–43, *Animal Liberation* (1975; repr., New York: Ecco, 2003), and "Morality, Reason, and the Rights of Animals," in *Primates and Philosophers*, ed. F. de Waal (Princeton, NJ: Princeton University Press, 2008), 140–58. The famine case is discussed in more detail below.

37. Judith Jarvis Thomson, "A Defense of Abortion," *Philosophy and Public Affairs* 1 (1971): 47–66. Psychologists have conducted experiments on subjects' responses to Thomson's violinist case and Singer's drowning child case. We discuss how moral consistency reasoning might be used to construct debunking arguments from this research in Kumar and Campbell, "On the Normative Significance of Experimental Moral Psychology."

complex cases depends on the degree of relevant similarity (or dissimilarity) that one can legitimately maintain between the targets of moral response.

Needless to say, reasoning about morals does not always involve consistency reasoning. Even when such reasoning is central, as it often is in arguments against racism or sexism, other considerations are also central, such as ignorance of nonmoral facts or conceptual confusions. A case in point is the change in moral response toward homosexuality discussed above. Change in religious belief may arise from nonmoral considerations, and that may itself undermine one's reasons for condemning gay sex.<sup>38</sup> But consistency reasoning can also play an important role, in considering similarities in the nonprocreative sources of pleasure between gay and straight sex. Once religious dogma is set to one side, what exactly would allow pleasure to be morally permissible in one case and not the other?

### III. INTEGRATED DUAL PROCESS MODEL OF CONSISTENCY REASONING

Thus far we have relied mainly on examples drawn from everyday life and moral philosophy in order to explain what consistency reasoning is and how it shapes our moral responses. In this section we provide a deeper understanding of consistency reasoning by characterizing it from within an MMDP model. The existence of moral consistency reasoning, besides providing a mechanism of moral change, entails that systems 1 and 2 do not merely run in parallel but regularly act in concert.

There is an obvious and tempting way of trying to understand consistency reasoning. Consistency reasoning, on this alternative, is strictly the application of our general capacity for reasoning in system 2. We use system 2 in order to resolve apparent inconsistencies in various descriptive, nonmoral domains, and consistency reasoning is simply one more such application. Obvious and tempting though this alternative picture may be, we believe it is wrong. Consistency reasoning is both ratiocinative and affective and issues from the joint efforts of the reasoning system and the intuitive system. This makes consistency reasoning distinct from other kinds of moral reasoning that are

38. Although a change in religious belief can also turn on consistency reasoning. See Richmond Campbell, "God, Evil, and Humanity," *Sophia* 23 (1984): 21–35.

applied to moral matters but not specialized for morality.<sup>39</sup> By locating consistency reasoning in the cooperative interaction between system 1 and system 2, we are able to illuminate the character of consistency reasoning as a distinctive form of moral reasoning and explain its power to shape moral thought and feeling.

Before we arrive at our dual process model of consistency reasoning, we want to offer some preliminary considerations that make our own view more attractive than the generalist alternative mentioned above, on which consistency reasoning is effected by system 2 alone. By showing how system 1—the primary job of which is to produce first-order moral judgments—can be enlisted in the second-order task of evaluating those judgments, our model shows how sophisticated moral behavior can emerge from more basic cognitive capacities.

When individuals engage in consistency reasoning, very often they cannot explain why a pair of judgments is consistent or inconsistent. This might be called a kind of second-order moral dumbfounding. Reactions to inconsistency are also quick, automatic, and they embody negative moral affect. One automatically feels moral disapprobation toward individuals, including oneself, when they exhibit moral inconsistency. Finally, inconsistency is established only when no morally relevant difference is found between the cases to which there are opposite moral responses, yet perceiving a difference as being morally relevant normally requires that system 1 be engaged. The reason is that a relevant difference is normally experienced emotionally and motivationally as morally relevant, and this experience can be the product only of system 1. All of this suggests that system 1 is at play in consistency reasoning, against the generalist alternative. Some of these claims are empirically tractable, and the experimental paradigms that guide other psychological research can be adapted in the service of testing our revised model. A sketch of how to go about testing our model experimentally against the generalist alternative will come at the end of the section.

We begin presenting our model by describing a form of reasoning that is different from consistency reasoning. Utilitarian cost-benefit reasoning, mentioned earlier, is an instance of a more general form of reasoning we shall call “moral principle reasoning,” or simply

39. Is consistency reasoning specific to the moral domain or, more broadly, the normative domain? We suspect that consistency reasoning is applied primarily to the moral domain but that it is also applied to other normative domains. Campbell's evolutionary account suggests that the “proper” domain of consistency reasoning is morality and that its “actual” domain includes other areas of normative thought and discourse (Richmond Campbell, “The Origin of Moral Reasons,” in *Logic, Ethics and All That Jazz: Essays in Honour of Jordan Howard Sobel*, ed. L. Johansson, J. Osterberg, and R. Sliwinski, Uppsala Philosophical Studies 57 [Uppsala: Philosophy Department, Uppsala University, 2009], 67–97).

“principle reasoning.” In principle reasoning we typically begin with a general moral principle, add certain empirical facts, and infer from these a more specific moral conclusion. The reasoning may be utilitarian if it begins with a utilitarian principle, but it may also be non-utilitarian if the principle concerns things like rights or justice. For example, we might reason as follows: everyone has the moral right to freely express an opinion so long as doing so does not result in grave danger to others; a certain law threatens to severely restrict our ability to express our opinions and does not prevent grave danger; therefore, the law is morally indefensible.

Principle reasoning takes place entirely in system 2. It is slow, conscious, controlled, effortful, and not intrinsically affective (although the inputs on which it operates may have affective etiologies).<sup>40</sup> Consistency reasoning is different. In consistency reasoning, we try to make our emotion-driven intuition about a “target” situation consistent with our emotion-driven intuition about a “base” situation, actual or hypothetical. In the course of this reasoning, we engage in conscious and deliberative processes, namely, in describing to others or ourselves various situations and identifying the most salient differences. However, our response to the target and base situations is quick, automatic, and affective. Moreover, we have a quick and automatic negative affective response to apparent inconsistency. Responses to apparent inconsistency, we hypothesize, are generated by the very same norms that produce our responses to the target and base situations.

First, we identify the salient differences between the two cases; second, these differences are fed as input to system 1; third, if none of the norms in system 1 are activated—if the difference is not perceived as morally relevant, thereby engaging system 1—system 1 issues in a negative affective response.<sup>41</sup> Once we have before ourselves the target situation, the base situation, the apparent inconsistency, and our affective responses to all of these, a motive to be consistent moves us to revise our moral responses in order to resolve the apparent inconsistency. Whichever response is less tenable, if any, is revised. Conscious, deliberative reasoning may reassert itself in trying to identify ways in which the target situation and the base situation are dissimilar in morally relevant ways. This can elicit new cases in which

40. Thus, determining whether free expression would result in “grave” danger or whether a law threatens “severe” restrictions on free expression would enlist system 1. But application of the principle to these inputs, to infer that the law is morally indefensible, requires only system 2.

41. Our description of inconsistency detection is sufficiently general so as to be neutral with respect to how moral norms are internally represented, whether by linguistically structured rules, connectionist prototypes, action-oriented schemata, etc.

the dissimilarity is erased, and the initial process is reengaged to check whether the inconsistency remains. The entire process, whether or not it is extended by consideration of potentially relevant dissimilarities, is markedly unlike principle reasoning because of the central role afforded to automatic, unconscious processing, in particular, our intuitive emotional response to apparent inconsistency. Figure 1, to which we will make further references below, gives a simplified representation of the different stages of this process.<sup>42</sup>

Recall the example of consistency reasoning concerning famine relief found in Singer's work.<sup>43</sup> What follows is a description of how this example fits our dual process model of consistency reasoning—and at the same time an explication of that model—although the same fit could likewise be described using other examples of consistency reasoning. We choose to focus on this example, first, because it is one in which the entire extended process of consistency reasoning is instantiated and, second, because it is an argument that has a demonstrable effect on some people's (although certainly not everyone's) moral attitudes and intuitions. Greene argues that moral theorizing among deontologists consists in blunt emotional responses and then elaborate post hoc rationalization.<sup>44</sup> We offer here a more friendly psychological characterization of a pattern of reasoning common to utilitarians and deontologists.

Singer is concerned with the fate of those starving in third world countries. This is the “target” situation. As it stands, although we think it would be good to help by donating money, we do not feel we have a pressing obligation to do so. Donation is, in other words, a matter of charity, not duty. However, Singer asks us to imagine ourselves walking along a shore and noticing that a young child is drowning in a shallow pond. We can save the child, but to do so would require ruining our new suit. This is the “base” situation. The response that Singer aims to evoke is virtually universal: we feel strongly that we have a pressing obligation to save the child, new suit be damned. But, as Singer argues, these two responses—no pressing obligation in the one case and a very pressing obligation in the other—appear inconsistent. In both cases our actions would save a life, and all we would have to give up is something of limited personal value or its monetary equivalent. What exactly would oblige us to help in one case but not in the other? Notice that although Singer is a utilitarian, the argu-

42. We offer this model of consistency reasoning as a working hypothesis. We are open to revisions of this model and open, too, to the possibility that different instances of consistency reasoning will recruit systems 1 and 2 in different ways.

43. Singer, “Famine, Affluence, and Morality.”

44. Greene, “Secret Joke of Kant's Soul.”

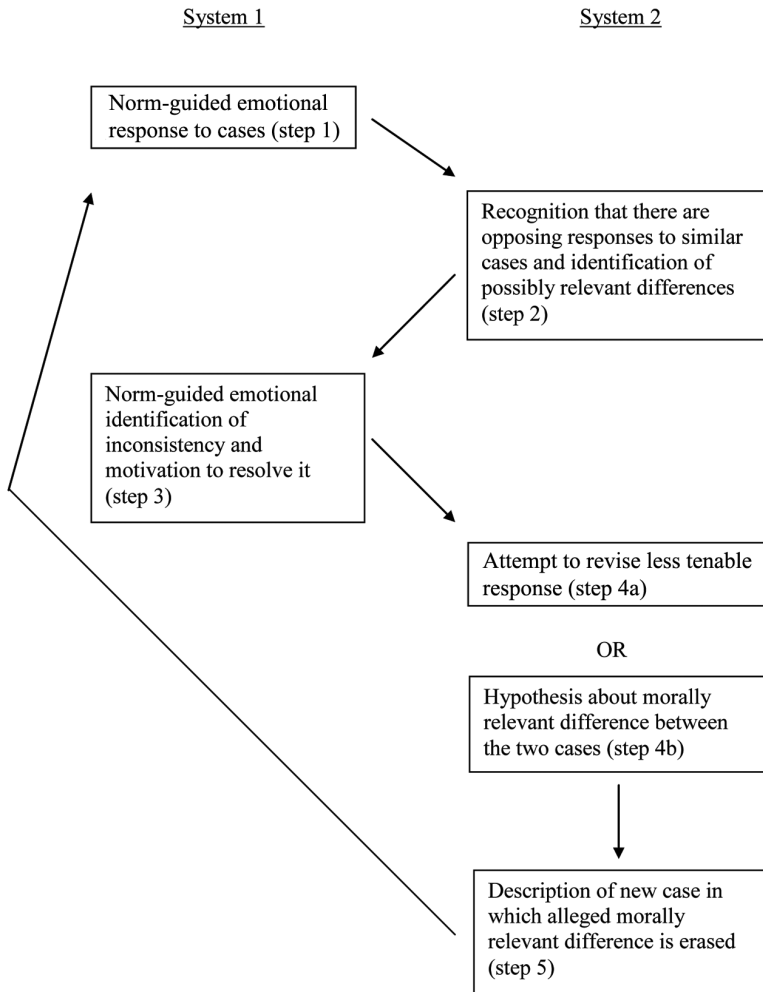


FIG. 1.—Dual Process model of consistency reasoning. Step 1: System 1 generates emotional responses to the target and base cases. Step 2: Consistency reasoning is engaged only if the cases are similar and the responses are opposed. Then system 2 identifies possibly relevant differences between the two cases, which differences are sent to system 1. Step 3: If the differentiating features do not activate any of the norms in system 1, a negative affective response (to our pair of responses) is generated. Step 4a: This leads to an attempt, in system 2, to revise the less tenable response. Step 4b: Alternatively, an attempt is made to identify a heretofore unseen relevant difference between the two cases. Step 5: In the latter case, this alleged difference can lead to description of new cases themselves subject to the demand for consistency.

ment does not appeal to a utilitarian principle. This, we believe, is part of what makes Singer's argument so powerful.

Our intuitive and reasoning systems both contribute to this piece of consistency reasoning. First of all, system 1 produces responses to the famine case (target situation) and the drowning child case (base situation; step 1, in fig. 1). Recognizing that the two responses are opposed and that the two cases are similar, system 2 identifies possibly relevant differences between the famine and drowning child cases, let's say, spatial proximity (step 2). This difference is fed as input to system 1. System 1 generates a negative affective response because the difference fails to activate the norms in system 1—that is, proximity is not a morally relevant feature according to those norms (step 3). Seeing that our responses to the two situations are apparently inconsistent and being motivated to resolve the inconsistency, we then reason that we must revise one or another of our judgments (step 4a).<sup>45</sup> Singer thinks, plausibly, that our response to the target situation is less tenable, and some people have agreed that we owe more to third world victims and have modified their behavior accordingly. Even among the majority of us who cannot entirely overcome selfish motivations, there is nevertheless often a partial but significant change in attitude toward famine relief.

Singer's example of consistency reasoning is especially instructive because some ethicists have attempted to show that his target and base situations are not relevantly similar and, hence, that the inconsistency between our moral responses is merely apparent (step 4b). For example, it has been argued that the social location of the child is a relevant difference; we have more reason to help the drowning child than starving third world children because we have special obligations to those in our community that we do not have to individuals in other communities.<sup>46</sup> The question of whether this response is apt need not detain us. For in any case it illustrates that an apparent inconsistency can be resolved by reengaging system 2 in an attempt to show that there really is no inconsistency—to describe ways in which the target and base situations are relevantly different—instead of revising one of our responses. Consistency reasoning can continue if new cases are described in which the allegedly relevant difference is erased and apparent inconsistency between our new pair of moral

45. To judge that there is an apparent inconsistency is to judge that there seems to be no morally relevant difference between the two cases. There may, however, turn out to be a hidden morally relevant difference.

46. Social location isn't the same thing as proximity. Individuals belonging to a particular community may find themselves in a distant place and, some people may believe, are more entitled to support from their home community than are people native to that place.

responses remains. Consistency reasoning is thus recursive, as indicated in figure 1.

To illustrate this last point, consider how Singer responds to an objection that claims there is a morally relevant difference between the target and base situations, namely, that you are the only one who can help the drowning child whereas there are millions of others who can aid those starving in the third world. Singer responds by introducing a new case that takes into account the allegedly morally relevant difference (step 5).<sup>47</sup> Suppose, Singer says, that there are others standing next to the pond doing nothing to help the drowning child. Intuitively, this does not lessen your obligation to help. So, Singer concludes, the difference in the original pair of cases of whether there are others who are in a position to help is not a morally relevant difference.

It is noteworthy that Singer articulates a general principle that is supposed to make clear the relevant moral similarity between the drowning child before one and the starving child who lives in another country: "If it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, then we ought, morally, to do it."<sup>48</sup> The objection just noted could be viewed as an objection to this principle—the principle is, allegedly, false when other people are also able to prevent something bad from happening—but Singer does not try to defend the principle directly. Instead he uses, as we have just noted, a further example to undercut the intuitive plausibility of the alleged morally relevant difference. The principle itself is not doing the moral work. Rather, its plausibility depends on whether the difference cited, that there are millions of others around to help the starving child, is credible as a morally relevant difference, and the way Singer tries to show that it is not is to give a brand new example, of many persons nearby ignoring the drowning child, where it will be obvious that the supposed difference is not morally relevant. It is, in short, consistency reasoning recursively applied that guides moral thinking in this instance, not deduction from a general principle.

The recursive nature of consistency reasoning helps clarify our disagreement with Haidt and Greene. Although we think consistency reasoning is operative in moral judgment, we do not deny that humans also have a tendency to engage in post hoc moral rationalization. This may be what Singer's critic does by citing the difference above. Recursion of consistency reasoning can help to avoid rationalization. By describing a new case in which other people are standing

47. Singer, "Famine, Affluence, and Morality," 232–33.

48. *Ibid.*



next to the pond, Singer shows that the counterargument cites a feature that is not morally relevant. As we have insisted, consistency reasoning is often a social process. Thus, another way in which we can avoid the trap of rationalizing is by listening to the arguments of others who, because of their different moral commitments, are not tempted to rationalize in the same way we are.

If we are right, there is a form of moral reasoning that, in contrast with principle reasoning, is specifically moral and engages both the reasoning system and the intuitive system. Before offering a sketch of how to empirically test the model, we wish to articulate three objections and our take on them.

First, critics may suggest that consistency reasoning is just a species of principle reasoning, but not one that calls on different moral principles. In consistency reasoning, on this suggestion, we reason by appeal to a single general principle to treat like cases alike, plug in our judgments about the target and base cases, and conclude that we must revise one or another of our judgments. This objection presupposes an inaccurate account of the kind of inconsistency to which we are responsive in consistency reasoning. As we have already made clear, our two responses are not by themselves inconsistent. There is no contradiction in judging that one is morally obliged to save a nearby drowning child but not obliged to save a distant starving child even if the costs are comparable. The problem is, rather, that the two responses together seem morally indefensible because we see no morally relevant difference between them.<sup>49</sup> The inconsistency arises out of our responses to the two cases in the context of our basic substantive moral commitments that inform us about what is morally relevant. Consider, by way of illustration of this point, the reaction of some religious conservatives to attempts to engage them in consistency reasoning. Some liberals have argued that current laws against gay marriage are no different from Jim Crow laws against interracial marriage. Many Americans, although condemning all such *de jure* racism, perceive morally relevant differences between the two situations because of internalized prohibitions against gay sex and sexuality. Thus, the

49. In this respect, consistency reasoning is also distinct from “universalization” reasoning described by R. M. Hare, who claims that moral judgments should be identical for cases sharing all universal properties: “Moral judgments are, I claim, universalizable in only one sense, namely that they entail identical judgments about all cases identical in their universal properties” (*Moral Thinking: Its Levels, Method, and Point* [Oxford: Clarendon, 1981], 108). Consistency reasoning is logically much more demanding, requiring identical judgments in cases where not all universal properties are shared, as Singer intends in his example of the drowning child. Hence, the normative demands of consistency reasoning, to be discussed in Sec. V, will be stronger.

inconsistency is morally substantive and cannot be simply an instance of deductive principle reasoning.

Second, one may argue that consistency reasoning might be a form of reasoning by analogy, where the latter is a domain-general kind of inductive reasoning.<sup>50</sup> Thus, if situations A and B are analogous in many respects and action type X is wrong in A, likely X is also wrong in B. Indeed, recognition of similarities between A and B might engage both system 1 and system 2. Consistency reasoning, however, is different because (as we have argued) it depends on morally substantive background considerations. In consistency reasoning, the inference that X is also wrong in B is permitted only if there is no morally relevant difference. There could be one even though A and B are very similar in other respects—or none, although A and B are very dissimilar in morally irrelevant respects (as in the King David example). The reasoning thus appears not to be domain general.

This leads us to a third and more serious objection. Critics may agree with our characterization of consistency reasoning as morally substantive but argue that it is implemented entirely in system 2. We consider our responses, reflect on our substantive moral commitments, and judge that they are together inconsistent. Although empirical tests of the sort described below are needed to assess this competing hypothesis, we are skeptical. We admit that consistency reasoning may sometimes be carried out in this sort of explicit way. However, this is not how the process is typically carried out, and we doubt it could be so carried out effectively in most cases. Instead of making explicit appeal to our substantive commitments, we have an automatic, intuitive, affective response to our judgments about the target and base cases. Just as we respond negatively to someone who fails to have the appropriate moral response to harm, we respond negatively to our inconsistent responses.<sup>51</sup> But their inconsistency is invisible to system 2. The negative response to our judgments issues from tacit commitments in system 1 that suggest the salient differences between the cases cannot be, in light of our tacit commitments, morally relevant—cannot justify our different responses. Nowhere in this process do we normally engage in the sort of fully explicit, step-by-step reasoning

50. Thanks to an anonymous editor for pressing this objection.

51. Our negative response to moral inconsistency is markedly different from dissonance reduction discussed in the psychological literature and by some philosophers, e.g., David Velleman, “From Self Psychology to Moral Philosophy,” in *Action and Freedom*, ed. James E. Tomberlin, Philosophical Perspectives 14 (Malden, MA: Blackwell, 2000), 349–77. There the agent is not aware of dissonance reduction, unlike the moral case in which the agent reacts consciously to “dissonance.” Moreover, the “dissonance” in the latter case arises, not in beliefs about ourselves but in moral beliefs and feelings regarding situations external to us.

our critic has in mind. In fact, we could not do so because normally recognition of inconsistency depends on the tacit commitments in system 1 for identifying morally relevant differences.

Whether our dual process model of consistency reasoning or a generalist model is correct must turn ultimately on which model is supported by empirical studies. We conclude this section by suggesting some possible experimental approaches.<sup>52</sup> First of all, one important prediction of our model is that consistency reasoning is domain specific. If subjects are better at detecting inconsistencies in the moral domain than analogous inconsistencies in the nonmoral domain, the best explanation may be that we have a system designed specially for morality that is employed in consistency reasoning. For example, psychologists have found that subjects more easily detect violations of deontic conditionals than violations of indicative conditionals, where the content of the conditionals is otherwise held constant.<sup>53</sup> Analogously, we should expect that subjects more easily detect inconsistencies among moral responses than other kinds of responses, holding the rest of the content constant.

Another set of studies would parallel work Haidt and Greene marshal in support of a dual process model of moral judgment. First, we might try to find experimental evidence of second-order dumbfounding, cases in which subjects make judgments of consistency or inconsistency but cannot cite the implicit norms on which they rely. Second, if consistency reasoning involves the emotion-driven system 1, manipulation of emotion should affect consistency reasoning as it has been shown to affect more ordinary moral judgments.<sup>54</sup> Third, fMRI studies may show activation of the areas of the brain associated with emotion when subjects engage in consistency reasoning. Finally, VMPC patients should manifest an impaired capacity for consistency reasoning. If predictions of this sort are borne out, generalist models of consistency reasoning might be ruled out.

52. An anonymous reviewer has cautioned that genuinely collaborative cross-disciplinary research between philosophers and experimental psychologists is required to devise appropriate empirical tests for our model of consistency reasoning. We agree and offer the following preliminary suggestions about testing with that in mind.

53. Leda Cosmides, "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task," *Cognition* 31 (1989): 187–276; Leda Cosmides and John Tooby, "Cognitive Adaptations for Social Exchange," in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, ed. J. Barkow, L. Cosmides, and J. Tooby (Oxford: Oxford University Press, 1992), 163–228.

54. See, e.g., Thalia Wheatley and Jonathan Haidt, "Hypnotically Induced Disgust Makes Moral Judgments More Severe," *Psychological Science* 16 (2005): 780–84; Jennifer Lerner, Julie Goldberg, and Philip Tetlock, "Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility," *Personality and Social Psychology Bulletin* 24 (1998): 563–74.

None of these tests, of course, would address moral emotional and motivational change over time as the result of consistency reasoning. Nor to our knowledge has there been experimental work on this topic. Longitudinal studies of the right kind would be difficult but not impossible to construct. Imagine a pool of subjects with moral responses to a type of situation, and imagine that these responses could be challenged on the basis of consistency reasoning. Supposing that a (large enough) subset of the pool would agree to discuss their views at designated intervals, these subjects might be randomly sorted into three groups. Members of the experimental group would be challenged during discussion at each interval by consistency reasoning. Members of one of the control groups would discuss their views at each interval with someone who disagreed with them but did not apply consistency reasoning, and those of the remaining group would discuss their views at each interval with someone who agreed with them. The groups would then be examined to see whether any moral change could be detected in their responses to the original situation over time.<sup>55</sup>

There are obvious difficulties in avoiding being misled by confounding variables, especially experimenter demand, but these might be mitigated sufficiently to make the results interesting, say by having the interviewers randomly switch roles several times over the course of the experiment. Also, lack of moral change among those subjected to consistency reasoning could be explained by their reaching consistent responses to various scenarios, or it could be, as sometimes happens, that subjects recognize that they are being inconsistent but do not care. The latter can happen when they distrust the person who has shown them to be inconsistent, for example, because they feel they have been tricked or fallen into confusion by accident rather than because of a problem in their moral responses. The task of designing measures to reduce the likelihood of these possibilities is difficult but could prove to be worth the effort.

#### IV. FUNCTIONAL ROLE OF CONSISTENCY REASONING

The functional role of the implicit norms that guide our moral-emotional responses is, roughly, to coordinate behavior and allow us to get along with one another. We take this claim to be uncontroversial, even if morality takes on additional meaning and significance once we begin to reflect on our moral commitments and identify with them. In this section, we rely on our empirical discussion above in

55. Researchers could adapt well-known techniques for measuring tacit responses, e.g., implicit association tests (e.g., Devine, "Stereotypes and Prejudice").

order to offer a characterization of moral consistency reasoning at the level of what it accomplishes. Consistency reasoning is a mechanism that functions to resolve practical conflict among our emotion-infused moral judgments.<sup>56</sup>

Must moral norms conflict? Famously, Kant and Mill each postulated a different single moral standard from which a consistent set of particular judgments or attitudes could be derived. Viewing morals from an empirical perspective, we find, however, a plurality of basic moral norms.<sup>57</sup> As such, they are bound to conflict to some degree, both within persons and between them. Such normative conflicts threaten to rob moral norms of the important benefits they provide. Moral consistency reasoning works to resolve moral conflict, both within and between persons.

These points were illustrated in Section II by the two historical examples of consistency reasoning. Let us now attempt to characterize the Baalsrud and King David examples in a general way. We begin with inconsistency in moral responses between persons—interpersonal inconsistency—using those examples. There are two kinds of interpersonal moral inconsistency that are illustrated in these examples, and each presupposes a context in which there exist one or more moral norms that have been socially internalized in most or at least many members of a group.

It will be helpful first to introduce some terminology. Moral norms, we suppose, guide moral responses (in emotion, motivation, belief, and behavior) to features of situations that prompt those responses. The features that would normally prompt the moral response are “reasons” for the responses of those who subscribe to the norm, in the sense of (*a*) being their causes and in the sense of (*b*) providing a partial justification for the response. Although there is much to be said about how features that cause moral responses become “reasons” for them in both senses *a* and *b*,<sup>58</sup> the concept of consistency reasoning is basic enough to require reference only to features that are “reasons” for moral responses in sense *a*. We concede that some might insist that moral norms by their nature demand responses to features that are reasons for moral responses in both senses *a* and *b*. These readers may wish, therefore, to read “moral norm” as referring only to protomoral norms. Nothing in the following analysis depends on

56. Campbell, “Origin of Moral Reasons,” offers an evolutionary explanation of this function, but the claim here does not depend on it.

57. See, e.g., Jonathan Haidt and Craig Joseph, “Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues,” *Daedalus* 133 (2004): 55–66.

58. For explanation of how natural selection could bring about the transition, see Campbell, “Origin of Moral Reasons.”

moral norms demanding responses that are justified, even partially, by reasons. "Reasons," as we use the term here, are simply the normal causes of the responses in those who are guided consciously or unconsciously by the norm in question. On the other hand, the following analysis does not exclude the more complex notion of reasons; it is simply neutral in this regard, allowing consistency reasoning to be applied with little moral sophistication.

(Case A) In the simplest case, one person does not respond morally as others do, where the others have internalized a moral norm requiring a certain response and thus disapprove of this divergence in which, given the same reason for response, someone fails to respond as they do or responds in an "incompatible" way. Responding in an incompatible way is responding in emotions and motivation in a way that would normally lead to doing something different from what normally would be done if the emotions and motivation were to follow the usual pattern. The example of King David's betrayal of Bathsheba's husband illustrates case A. Most of those with whom King David would be apt to discuss moral matters would respond to the prospect of such betrayal of trust with revulsion and condemnation, even if they would be afraid to express it openly. King David did not respond in this way. He allowed himself to act on his lust for Bathsheba and thus to betray the trust of his loyal general. His is a response incompatible with that demanded by the moral norm that calls for honoring such trust and condemning its betrayal.

(Case B) The more complex case obtains when two moral norms are internalized and expressed in many instances without conflict until a new situation arises in which the norms cannot be applied without inconsistency. Such conflicts are commonplace in moral life. Keeping a promise to do X for the reason that one has promised to do X is consistent in most cases with avoiding harm, yet circumstances can arise in which to do X, which one has promised, will result in harm in these circumstances. Moreover, the gravity of the harm may mean that emotion and motivation leading toward the avoidance of harm in this case are equal but opposite to those leading toward keeping the promise. Since both patterns of moral response are internalized, each person will respond morally with disapproval if the promise is broken but will also respond morally with disapproval if the harm is done, unless there is a way to remove the moral inconsistency. The inconsistency is interpersonal because the combination of emotion and motivation in one person that leads toward keeping the promise is incompatible with the combination of emotion and motivation in another person that leads toward avoiding harm. The inconsistency in case B is also obviously intrapersonal since the warring combinations exist in each person. The Baalsrud example illustrates case B:

Marius's challenge forced his mother to face a conflict weighing the norms of beneficence and loyalty against the norm not to risk grave harm.

Since internalization of moral norms varies somewhat (sometimes greatly) from person to person, case A types of interpersonal inconsistency are to be expected to some degree, but their frequency often can be reduced through better social training. In particular, greater interpersonal consistency can be created through stronger rewards for consistency in normative response and stronger punishments for lack of it. In this respect, case B kinds of inconsistency are importantly different. No matter how strong or how effective the training is to conform to the norms in question, the inconsistency in case B will remain for situations in which the norms become incompatible. The reason is simply that in case B conformity to the norms can be complete in other contexts, involving no inconsistency. The inconsistency arises because the two separate normative patterns are brought into conflict by new situations in which they cannot both be realized. Another social mechanism is needed to address this problem if situations of conflict are frequent enough to undermine allegiance to the respective norms and thereby destroy their benefits.

We see moral consistency reasoning as a device that can address both kinds of cases. Interestingly, it is a social device since it typically (but not necessarily) operates in a social setting in which reasons for individual responses are questioned, but while the operation of the device is typically social, its immediate target is intrapersonal inconsistency. That is, the social pressure exerted in consistency reasoning is directed at a single individual's inconsistency in moral responses. This pressure, however, functions to remove the interpersonal inconsistency, whether the latter is case A or B, and thus, when working properly, the device is able to resolve the problem. Notice that both kinds of inconsistency are defined by reference to moral norms, but consistency reasoning focuses on the responses to reasons in a single individual. In particular, when an individual gives incompatible moral responses in situations where the reasons prompting the responses are of the same type, the individual exhibits (intrapersonal) moral inconsistency and can be a target of consistency reasoning. Consistency reasoning is initially the process of exposing inconsistency in an individual's actual or potential moral responses to reasons through, for example, (*a*) asking for reasons for moral responses in a given situation and then (*b*) asking for responses (especially approval or disapproval) to other actual or possible situations that contain similar reasons for response. This process is well illustrated in the two examples above where the combination of *a* and *b* result in exposing intrapersonal inconsistency.

It is worth noting that one can restore consistency in one's responses to reasons in case A only by conforming again to the general pattern of others since the cost of restoring personal consistency by rejecting all conformity to the general pattern of others has the same consequence as remaining interpersonally inconsistent. Although one who completely rejected conformity to the general pattern could be predictable (if one were believable, one would be expected never to conform), one would remain unreliable in that respect for cooperation in gaining the mutual benefits of conforming to the general pattern. Importantly, case B is different since consistency cannot be restored through simply conforming to the general pattern of others. However, consistency can be restored in case B by subtly modifying the prompting reasons in either of the conflicting patterns so that the old patterns remain the same except for the new situations, and there the modification of one of the prompting reasons removes the conflict. The high probability of death to oneself and one's community if an attempt is made to save one person lost its power in Marius's family to prompt sacrificing the one for the many when the family identified with the vulnerability of the stranger as morally indistinguishable from its own in the face of a vicious common enemy.

Obviously, a problem remains since it is unclear in general which prompting reason needs modification or how it is to be modified. The problem of choice finds resolution only in a social search for modification in reasons that may resolve inconsistency without introducing new cases of inconsistency. Case B inconsistencies need not have any clear resolution, however. Indeed, case B inconsistency may be hard to resolve through the social process of consistency reasoning. There are many examples of this difficulty in modern society when normative conflicts are deeply entrenched. Nevertheless, what happens often enough is that the consistency reasoning is carried out in a context of many other patterns of norm-guided response. The pressure for intrapersonal consistency will be likely to bear on many tempting but ultimately unsatisfactory attempts at restoring intrapersonal consistency. In short, the means of gaining consistency in the given case generates inconsistency with other moral norms that are already firmly established in the group, thus affording a resolution in which responses to the new situation become consistent implicitly with already established patterns. There is moral innovation in responding differently in the new cases but innovation that is minimally disruptive in its implications.<sup>59</sup>

59. Richmond Campbell and Jennifer Woodrow suggest how such social resolution of normative conflict about new cases through a basic form of recursive consistency reasoning might have evolved by natural selection ("Why Moore's Open Question Is Open: The Evolution of Moral Supervenience," *Journal of Value Inquiry* 37 [2003]: 353–72).



Fundamental differences exist between the two examples. As noted, the biblical example is case A since King David's moral responses are inconsistent with the majority, but in the decision of Marius's mother, the community is initially in conflict with itself, illustrating case B. Nevertheless, the style of consistency reasoning used to bring the community into moral consistency is basically the same since in each case the exposure of inconsistency in a pattern of moral responses to types of situation forces individuals, in the context of entrenched patterns of moral response, to modify their moral responses on pain of inconsistency. Each change takes time, by consideration of hypothetical but similar cases posed by interlocutors, but the moral change in the individuals is effected through consistency reasoning. In case A the change would not count as a change within the moral values of the majority; in case B the moral change has wider scope and is innovative since the community moves to a new moral perspective on how to address its mutual dilemma.

## V. PHILOSOPHICAL SIGNIFICANCE

We have sought to provide an improved dual process model of moral cognition by focusing on the psychological role of consistency reasoning in moral thinking, as well as the social role of consistency reasoning in securing the cooperative functions of morality. Our aim has been to give a more adequate model of how humans think morally rather than how they morally ought to think. Our theory is intended, in other words, to be descriptive rather than normative in its function. Nevertheless, the theory can be instructive when put in the context of normative theory.

Consider the celebrated normative model of reflective equilibrium (RE), originally advanced by Goodman as a nonfoundational method of assessing principles of deduction and induction and applied famously by Rawls to assess principles of justice.<sup>60</sup> While the idea of RE was advanced in each case as a normative standard for assessing normative principles, it is possible to transform the descriptive model of consistency reasoning (CR) into a parallel normative standard for assessing particular normative judgments because of two striking structural similarities between them. First, both RE and CR concern themselves with moral inconsistency. RE concerns inconsistency between a

60. Nelson Goodman, *Fact, Fiction, and Forecast* (Cambridge, MA: Harvard University Press, 1955), 65–68. John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), 20–21, 48–51. An anonymous editor has suggested that Rawls did not think of inconsistency in the RE model as always involving a general principle. Perhaps, but our aim is philosophical rather than exegetical: to make clear important limitations of the principle-based model of moral inconsistency.

general moral principle and a particular moral judgment, and CR concerns inconsistency between two particular moral judgments. Still, in both cases, inconsistency is a justifiable reason to resolve the conflict. Second, the justification for resolving either conflict by rejecting one judgment (or principle) in favor of another is relative to implicit background norms that are justified or that, in the case of CR, reasoners assume are justified.

We have been at pains to make the second point clear for CR in two related ways. (a) Genuine inconsistency between opposite moral responses to similar cases obtains only if there is no morally relevant difference between the cases, and judging if the latter is so depends on implicit moral norms. (b) Once inconsistency is judged to obtain, however, there is the further issue of determining which of the responses in conflict is less tenable, and that too, as we have already emphasized, depends on still other moral norms. In both *a* and *b* the person or persons engaged in CR act as if the implicit moral norms are justified. Although the role of background norms does not change the descriptive nature of CR, this structural similarity between CR and RE is critical for transforming CR into a normative standard, as we shall see in a moment.

Understanding the parallel role of background moral norms in RE requires that we sketch the idea of “wide reflective equilibrium” (WRE), a sophisticated version of RE that has been widely accepted and reveals the complex nature of inconsistency on this model.<sup>61</sup> A set of moral principles *P*, a set of particular moral judgments *J*, and a set of general background theories *T* can come into logical conflict, when *P* and *T*, together with empirical data, entail a normative judgment *J*\* that is logically inconsistent with *J*. The principles and judgments *P* and *J* are moral in their content, while *T* contains moral and nonmoral theory. On pain of being inconsistent in one’s overall moral perspective, change must occur somewhere to restore consistency. *P* might be modified or *J* or even *T*. Since WRE is nonfoundational, no member of this triad is beyond criticism or rejection. At times *J* may be so deeply entrenched that either *P* or *T* will have to be modified; the theoretical virtues of *P* and *T* may be at other times strong enough that *J* should give way.

Although this simple sketch vastly oversimplifies the complexity and power of WRE, the basic idea is familiar and attractive enough for us to proceed to complete the comparison with CR regarding the role of background norms. Just as inconsistency in CR depends on implicit moral norms (about morally relevant differences), conflict

61. Norman Daniels, “Wide Reflective Equilibrium and Theory Acceptance in Ethics,” *Journal of Philosophy* 76 (1979): 256–82.

between P and J can depend on moral norms contained in the background theories T. Moreover, how an inconsistency is to be resolved will be itself a moral matter, just as it is for those who are engaged in CR. For example, if a judgment in J is rejected in favor of P and T, this decision may be justified because P and T together with empirical data entail another set of moral judgments (including others in J) that are regarded independently as being acceptable. In sum, both WRE and CR are about the resolution of inconsistency, and the resolution is justified (in WRE) or assumed to be justified (in CR) only relative to other moral norms that are justified (WRE) or assumed to be justified (CR).

This relativity to background moral norms makes clear the normative status of WRE but also reveals how CR may be transformed into a new standard of evaluation with comparable normative status. Suppose in applying WRE we find that in light of empirical data some of our moral principles and judgments are inconsistent relative to some background moral norms in T. (1) If these background moral norms are justified, then the inconsistency in moral perspective is undesirable and should be resolved. We take this normative assumption to be uncontroversial, although soon we will support it with theoretical considerations. How the conflict should be resolved, as we have already noted, depends on still other moral considerations (as observed in the last paragraph), but (2) if they are justified, then the resolution of the inconsistency that accords with those moral considerations is likewise justified. We take this second normative assumption to be as uncontroversial as the first.

Given parallel assumptions, CR can now easily be transformed into a standard of justification with comparable normative status. Inconsistency between particular moral judgments in CR is undesirable and should be resolved. This claim is true, of course, only if the inconsistency determined in CR is genuine, that is, only if there really is no morally relevant difference between the cases in question. The latter turns on whether the implicit moral norms bearing on morally relevant differences are really justified, not merely assumed to be justified from the standpoint of those engaged in CR. Let NCR, then, be a standard for justified resolution implying that (1) inconsistencies between particular moral judgments should be resolved when they are genuine inconsistencies according to justified norms about morally relevant difference and (2) in such cases the less tenable of the moral judgments, given justified norms that determine their tenability, should be rejected. We now have a normative standard NCR (modeled on CR) that is as credible in normative status as WRE. We have done this simply by making the normative stipulation that the implicit

moral norms on which determination of inconsistency and justified resolution depend are themselves justified.<sup>62</sup>

A fundamental question arises at this juncture: Does the normative standard of NCR contain anything of normative significance not already contained in WRE? We will defend an affirmative answer, and our argument for it will turn in large part on whether our empirical theory of CR is sound, at least in broad outline. In short, our descriptive model of CR is normatively significant because its normative counterpart NCR goes beyond WRE to add a new normative perspective to understanding moral inconsistency. The argument has two parts. First, we show how NCR covers new moral ground because the source of the inconsistency is different in WRE and NCR; second, we show how the nature of the inconsistency is different for each standard and why that is important for understanding different kinds of moral change.

The first part is relatively straightforward, given the points already made. The source of inconsistency in WRE, but not in NCR, is the conscious and explicit conflict between a general moral principle and a particular moral judgment (in the context of explicit background assumptions, as noted earlier). The source in NCR is instead the conscious and explicit consideration of opposing moral judgments about two similar cases (with the background assumption that the two cases are similar in morally relevant respects). The first kind of conflict may be illustrated in the apparent inconsistency between classical utilitarianism and the moral judgment that the institution of slavery is unacceptable (allowing economic and psychological assumptions that would enable one to deduce the contrary moral judgment from the utilitarian principle). By contrast, the opposing moral judgments about the relevantly similar situations of Jan and Marius manifestly do not comprise a conflict between one of these judgments and a general principle. It is possible to tell a different story in which Marius's mother, to defend her view that the family should not attempt to rescue Jan, invokes a complex principle (e.g., that weighs unintended but foreseen harm to many against the deliberate sacrifice of a compatriot to save the many). She did not do that, and most people would not. Moreover, this kind of conflict is not unusual in its focus on particular judgments. Such cases, as we have tried to indicate, abound in everyday life.

62. We could similarly transform WRE into a descriptive model (DWRE) by treating the implicit moral norms for determining inconsistency and its resolution as norms assumed to be justified by those who see themselves as endorsing a moral principle that conflicts with a particular moral judgment of theirs and who are motivated to resolve the conflict. DWRE would predict a move toward WRE.

That such cases are common is exactly what we would expect given the MMDP model that we defended and elaborated in earlier sections. MMDP does not exclude the formation of moral principles and deductions of particular moral claims based on those principles and other assumptions (activities often displayed in professional philosophical journals). The latter activities engage primarily system 2. However, given MMDP, we would expect that conflicts in moral judgments would often arise when system 1 yields opposite responses to situations that are distinct but similar enough to invite the question of whether there are one or more morally relevant differences. We take as beyond serious debate that the expected pattern obtains in countless instances. MMDP plausibly explains why.

Even when a philosopher invokes a universal principle, as Singer did to reinforce his argument that one should save a child in another country, it is often necessary, in order to defend the principle in the face of an objection, to solicit another moral judgment about a distinct but relevantly similar example that is in conflict with the first judgment. WRE does nothing to address this kind of inconsistency since no universal moral principle is being used to address the objection. The conflict in the latter case is between two particular judgments. A different standard, one that applies to particular judgments in conflict, would be appropriate, and NCR fits the bill. That is not to say that NCR needs to be consciously applied. It is enough, in Singer's example, that it is completely obvious that one should still save the drowning child, at almost no cost, when others can do likewise but won't. It would be gratuitous to add that the implicit norms that account for this judgment are justified. That is assumed by all who find Singer's rebuttal convincing.

The nature of the inconsistency that is normally at work in examples of CR adds a further dimension to the normative significance of NCR. Normally what is experienced by someone who has opposite moral responses to morally similar examples is practical inconsistency in emotions and motivation. Marius's mother was initially morally repelled by the extreme danger that a rescue attempt would pose for her family and community and was morally motivated to discourage any rescue attempt. When her son posed the Oslo scenario, however, she also found herself similarly opposed to the idea of Marius being used as a means to appease the Nazis. Unable to see a morally relevant difference between the cases, she found her moral feelings and motives to be in conflict. Although the Oslo example is hypothetical, the dilemma for her wasn't merely theoretical. We can reasonably infer that her sense of herself as a morally responsible person in her family and community would have been in jeopardy if the tension were left unresolved. How would she know how to feel or act toward Jan? How

could those close to her know if they could rely on her if they decide to help Jan? Her moral confusion would be practical in these ways and others because it directly affects her emotionally and motivationally in her immediate relationships. As it turned out, she perceived her response to the Oslo example as more tenable, and she began to perceive Jan as a Norwegian boy like her son, someone needing and deserving their aid. Her resolution was at once both moral and practical.

This understanding of the resolution fits well with the MMDP model. Not only are the conflicting judgments normally experienced emotionally and motivationally, but also the inconsistency itself is normally felt to be uncomfortable. There is a cluster of reasons why such inconsistency is often felt to be uncomfortable and needing resolution, as noted in the preceding section. This kind of moral confusion makes one appear unreliable and untrustworthy, especially in groups whose members see themselves as strongly interdependent, because it serves to undermine the efforts of others to coordinate their activities around one; whatever benefits normally accrue to cooperation on moral issues would be undermined, at least temporarily, because without resolution moral choices in this regard become unpredictable. Such confusion, however, has wider implications. Being perceived as unreliable is not good for the person who is confused, and that is normally a motive for change, but more generally the benefits of stable moral expectations for the well-being of the group are also threatened. Groups tend not to cope well morally when moral confusions affect their members, unless they can resolve them. CR, as we have argued, is a social mechanism that functions to resolve moral inconsistency.

On the other hand, moral confusion and conflict have their place.<sup>63</sup> Significant moral change is unlikely to occur without moral conflict, and justified moral changes in such cases, for example, when unjust but entrenched moral norms are challenged, requires a justified resolution of the moral confusion and conflict. From a normative perspective, then, the practical dimension of this kind of moral inconsistency fits well with NCR since it explains how a justified moral resolution of such conflict conforming to NCR can motivate change in the moral functioning of the community in which it occurs.

It is striking that the social and practical dimensions of moral inconsistency are absent from WRE. Indeed, in Rawls's original discussion of RE the issue of conflicts between principles and considered

63. We thank an anonymous editor for pressing us on this point.

judgments is abstracted away from any individual or group,<sup>64</sup> and subsequent elaborations have continued to represent the inconsistency as a logical relationship between beliefs contemplated abstractly. Perhaps WRE could be formulated as a way of resolving practical conflict, say between the emotional and motivational commitment to a universal moral principle and a similarly embodied moral judgment about a particular case. That is not, however, how WRE has been conceived as a moral standard for the justification of universal principles and particular judgments. Moreover, MMDP suggests that the formulation of principles is carried out deliberatively and without affect by system 2. If someone were to claim that our commitment to principles regularly carries with it emotions that would generate practical inconsistencies, that would require special defense. Thus, we conclude that NCR has normative significance that cannot be reduced to or explained by the normative force of WRE.

We consider in closing the following objection. WRE entails that moral judgments must be capable of entering in deductive logical relations or at least mimic that capacity,<sup>65</sup> for that is the means by which inconsistency is discovered. However, it is not clear that moral judgments, as conceived in the NCR model of moral reasoning, must enter into logical relations or even mimic doing so. After all, most moral judgments on the latter model are conceived as products of the emotion-driven, unconscious, intuitive system 1. Although Rawls presented the RE model as neutral on the metaethical issue of how to interpret the content of moral claims, the NCR model appears to fit best with an expressivist conception of moral judgments as thoroughly noncognitive, emotional states. It might be possible to render their content in a form that at least mimics true-or-false propositions, but it is far from obvious how to do that. Thus, there appears to be a tension between NCR and WRE. Can the two ways of thinking about change in moral judgment work together?

In several other places, we defend a “hybrid” conception of moral judgment.<sup>66</sup> Moral judgments are neither primarily true-or-false moral

64. “I shall not even ask whether the principles that characterize one person’s considered judgments are the same as those that characterize another’s” (Rawls, *Theory of Justice*, 50).

65. For discussion of the latter, see, e.g., Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press, 1990).

66. Richmond Campbell, *Illusions of Paradox* (Lanham, MD: Rowman & Littlefield, 1998), 169–75, “The Problem of Moral Judgment,” in *Engaged Philosophy: Essays in Honour of David Braybrooke*, ed. S. Sherwin and P. Schotch (Toronto: University of Toronto Press, 2007), 249–72, and “What Is Moral Judgment?” *Journal of Philosophy* 104 (2007): 321–49; Victor Kumar, “Reduction of Moral Judgment” (unpublished manuscript, University of Arizona, 2011).

beliefs nor primarily states of moral motivation and emotion. We argue instead that moral judgments normally are constituted by both elements, and the elements are related differently than they are in internalist and externalist forms of moral cognitivism.<sup>67</sup> Unlike other hybrid theories, ours allows either element to dominate in the process of moral transition exemplified in earlier examples. Either a combination of moral emotion and moral motivation can initiate moral change or moral belief can. For example, in the change in moral perspective on homosexuality discussed above, a new moral belief arrived at through moral reflection may not initially be linked to new moral feelings and desires, or new moral feelings and desires triggered by new examples and experiences, such as arising from being the target of discrimination, may initiate change in moral perspective without immediate change in moral belief.<sup>68</sup> We suggest that such a hybrid conception of moral judgment allows moral judgments to play a dual role in moral change when inconsistencies of different kinds are found in one's moral outlook, since moral beliefs allow deductions of the kind countenanced by WRE and emotional and motivational responses to new situations allow reasoning of the kind exemplified in NCR.<sup>69</sup> This conception of moral judgment neatly fits the integrated dual process model of moral cognition in which the interaction of the two systems allows complex moral change in beliefs, feelings, and desires that is integral to CR.<sup>70</sup>

More precisely, inconsistency in beliefs involving belief in a general moral principle and a moral belief about a particular situation to which the principle applies (in the context of other beliefs) is a justification for change in belief on the WRE model. At the same time, incompatibility in moral emotion and motivation between two moral judgments about particular cases for which there is no morally relevant difference is a justification for change in moral emotion and motivation on the NCR model. On the same model, modified to apply also to moral beliefs, incompatibility of moral beliefs between two

67. Others have developed hybrid theories in which either the cognitive element (David Copp, "Realist-Expressivism: A Neglected Option for Moral Realism," *Social Philosophy and Policy* 18 [2001]: 1–43) or the motivation element (Michael Ridge, "Ecumenical Expressivism: Finessing Frege," *Ethics* 116 [2006]: 302–36) is more central to moral judgment. Our view is more robustly hybrid.

68. For application of the hybrid theory to these examples, see Campbell, "What Is Moral Judgment?"

69. For a unified normative model with WRE as a special case, see Richmond Campbell, "Naturalizing Moral Inconsistency: Two Models and a New Direction" (unpublished manuscript, Dalhousie University, 2011).

70. The argument that a dual process model supports a hybrid conception of moral judgment is developed in more detail in Kumar, "Reduction of Moral Judgment."



moral judgments, again about distinct cases between which there is no morally relevant difference, is a justification for change in moral belief. In sum, we have in the combination of WRE and NCR, once revised to accommodate moral beliefs and moral emotion and motivation, an expanded normative model for justifying change in moral judgment.<sup>71</sup>

## VI. CONCLUSION

Our primary aim has been to improve the dual process model of moral cognition that is gaining increasing acceptance among psychologists concerned with moral cognition. While we believe that the dual process view of moral thinking in the minimal form of MMDP is correct in its understanding of moral cognition as a product of two fundamentally different neurocognitive functions, we take issue with the associated view that moral reasoning, when it does not involve calculating costs and benefits, is typically nothing more than post hoc rationalization. We not only raise problems for the latter hypothesis (noting, e.g., that the experiments offered to confirm it are structured synchronically rather than diachronically) but also offer a revised conception of the dual process model in which the two systems work interdependently when subjects engage in moral consistency reasoning.

By focusing a philosophical perspective on this line of empirical research, we hope that we may prompt new avenues of experimental research on moral reasoning, especially reasoning that leads to moral change. We hope to have shown, too, how an empirical approach to moral cognition suggests promising new directions in normative theory that may in turn again redirect empirical research.

71. Here and throughout, we mean justifying *some* change rather than a particular change. Which particular change to create consistency would be justified, as we have noted earlier, depends on the relative strength of one's substantive moral commitments and whether they are themselves justified.