



Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

FlashReport

Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie

Joshua D. Greene*

Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138, United States

ARTICLE INFO

Article history:

Received 11 January 2009

Available online xxxx

Keywords:

Moral judgment

Moral psychology

Dual-process

ABSTRACT

A substantial body of research supports a dual-process theory of moral judgment, according to which characteristically deontological judgments are driven by automatic emotional responses, while characteristically utilitarian judgments are driven by controlled cognitive processes. This theory was initially supported by neuroimaging and reaction time (RT) data. McGuire et al. have reanalyzed these initial RT data and claim that, in light of their findings, the dual-process theory of moral judgment and the personal/impersonal distinction now lack support. While McGuire and colleagues have convincingly overturned Greene et al.'s interpretation of their original RT data, their claim that the dual-process theory now lacks support overstates the implications of their findings. McGuire and colleagues ignore the results of several more recent behavioral studies, including the study that bears most directly on their critique. They dismiss without adequate justification the results of a more recent neuroimaging study, three more recent patient studies, and an emotion-induction study. Their broader critique is based largely on their conflation of the dual-process theory with the personal/impersonal distinction, which are independent.

© 2009 Elsevier Inc. All rights reserved.

My collaborators and I have developed a dual-process theory of moral judgment (Greene, 2007a; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001), according to which characteristically deontological judgments (e.g. disapproving of killing one person to save several others) are driven by automatic emotional responses, while characteristically utilitarian judgments (e.g. approving of killing one to save several others) are driven by controlled cognitive processes. This line of research was inspired by a philosophical puzzle known as the Trolley Problem (Fischer & Ravizza, 1992; Thomson, 1985): In response to the *switch* dilemma (previously referred to as the *trolley* dilemma), people typically judge that it is morally acceptable to divert a runaway trolley that threatens five lives onto a side track, where it will run over and kill only one person instead (Greene et al., 2001; Mikhail, 2000; Petrinovich, O'Neill, & Jorgensen, 1993). In response to the contrasting *footbridge* dilemma, people typically judge that it is morally unacceptable to push someone off a footbridge and into the path of a speeding trolley, saving five people further down the track, but killing the person pushed. The "Problem" is to explain why people respond (or ought to respond) differently to these two dilemmas.

In studying these dilemmas, our primary aim was to better understand the respective roles of emotional/automatic vs. con-

trolled cognitive processes in moral judgment. More specifically, we aimed to test our dual-process theory by collecting functional magnetic resonance imaging (fMRI) and reaction time (RT) data to test the following two claims: (1) People's characteristically deontological disapproval of actions like the one proposed in the *footbridge* dilemma are driven by automatic negative emotional responses. (2) Utilitarian approval of harmful actions is driven by controlled cognitive processes. (Utilitarian judgments occur often in response to dilemmas like the *switch* dilemma and less frequently in response to dilemmas like the *footbridge* dilemma.) Our secondary aim was to propose a preliminary theory concerning the features of the *switch* and *footbridge* dilemmas that cause people to respond so differently to them. This secondary aim was foisted upon us by the technical requirements of fMRI. We could not simply examine the *switch* and *footbridge* dilemmas in isolation because fMRI data are too noisy. Instead we had to develop two sets of dilemmas, one with the relevant features of the *switch* dilemma and one with the relevant features of the *footbridge* dilemma. We did not know which features were the relevant ones, but we hazarded a guess, which became the "personal/impersonal" distinction. Dilemmas, like the *footbridge* dilemma, in which the action would cause (a) serious bodily harm, (b) to a particular person or group, where (c) the harm does not result from deflecting an existing threat, were classified as "personal." The rest were classified as "impersonal." We were aware of problems with more familiar distinctions from the philosophical literature on the Trolley

* Fax: +1 617 4953898.

E-mail address: jgreene@wjh.harvard.edu

Problem (Fischer & Ravizza, 1992), such as the distinction between intended and foreseen harm (Thomson, 1985), and expected that our personal/impersonal distinction would soon be replaced or substantially revised (Greene et al., 2001).

McGuire and colleagues (this issue) reanalyzed the RT data from Greene et al. (2001), and their findings do indeed undermine our original interpretation of those data. We reported that judgments approving of “personal” harmful actions took longer than judgments disapproving of those actions. Because such approval is generally motivated by utilitarian considerations (saving more lives), we interpreted these results as supporting our claim that utilitarian judgments are driven by controlled cognitive processes, the engagement of which is reflected in longer RTs. McGuire and colleagues have shown that the effect we reported is an artifact: In the subset of dilemmas in which there is a genuine conflict between utilitarian considerations and other considerations (as in the *footbridge* dilemma), there is no RT effect. The apparent RT effect was generated by the inclusion of several “dilemmas” in which a personal harm has no compelling utilitarian rationale. These dilemmas reliably elicited fast, disapproving judgments, skewing the data.

McGuire and colleagues’ reanalysis is an excellent piece of scientific detective-work, and it serves as a lesson to me and, I hope, other researchers. However, their critique dramatically overstates the implications of their findings for the dual-process theory of moral judgment. Their critique has two principal problems: First, it unjustifiably dismisses and ignores more recent research supporting the dual-process theory, research that avoids the methodological problem they have identified. Second, it conflates two different scientific ideas: the dual-process theory of moral judgment and the personal/impersonal distinction as drawn in Greene et al. (2001). This conflation leads them to mischaracterize their own critique and is related to their unjustified dismissal of more recent evidence.

First, we will consider the evidence that McGuire and colleagues ignore. The problem identified by McGuire and colleagues was first brought to my attention by Liane Young (personal communication) who performed a similar reanalysis of our 2001 RT data. Prompted in part by her discovery, my colleagues and I conducted a cognitive load study (Greene et al., 2008) aimed at generating stronger evidence for the implication of controlled cognitive processes in utilitarian moral judgment. This study focused on “high-conflict” personal moral dilemmas (Koenigs et al., 2007) that (a) propose a harmful action with a clear utilitarian rationale and (b) reliably elicit conflicting judgments from normal participants. (The *footbridge* dilemma is a high-conflict dilemma, but other dilemmas more reliably elicit disagreement among subjects.) Subjects responded to these dilemmas under cognitive load and in a control condition. The load selectively interfered with the utilitarian judgments, increasing their RTs, but had no effect on RT for the deontological judgments. (The RTs for the deontological judgments were non-significantly *faster* under load.) These results more effectively make the point we attempted to make with our original RT data: Utilitarian judgments depend preferentially on controlled cognitive processes (which are susceptible to interference by cognitive load). I emphasize that these results in no way depend on the personal/impersonal distinction, as “personal” and “impersonal” dilemmas were never compared in this study. Nor do these results depend on data from the “low-conflict” “personal” dilemmas that artificially generated the RT effect in Greene et al. (2001). Finally, I note that the selective effect of load on utilitarian judgment was also observed in an item-based analysis.

Next we turn to McGuire et al.’s conflation of the dual-process theory and the personal/impersonal distinction. According

to the dual-process theory, people respond negatively to the *footbridge* dilemma because something about the action in this dilemma elicits a prepotent negative emotional response, one that is not elicited by the action in the *switch* dilemma, at least not as strongly. This negative emotional response conflicts with (and typically out-competes) the controlled cognitive processes that favor utilitarian judgment in this case. Note that this theory, as stated, says nothing about *why* the *footbridge* dilemma elicits a stronger negative emotional response than the *switch* dilemma. It could be because the harm in that case is more “personal” as defined in Greene et al. (2001), because it’s intentional (Cushman, Young, & Hauser, 2006; Mikhail, 2000; Moore, Clark, & Kane, 2008; Schaich Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006), because it involves an intervention on the victim (Waldmann & Dieterich, 2007), because it is more direct (Moore et al., 2008; Royzman & Baron, 2002), because it involves physical contact (Cushman et al., 2006), because it involves a combination of “personal force” and intention (Greene, Cushman, Stewart, Lowenberg, Nystrom, & Cohen, in press), or for some other reason. In other words, the dual-process theory could be completely right, even if the personal/impersonal distinction is completely wrong. The reverse is also true. The computations attributed to distinct systems by the dual-process theory could, in principle, be accomplished by a single system employing a weighted combination of Greene et al.’s (2001) three “personalness” criteria and a utilitarian principle.

McGuire and colleagues emphasize their doubts about the personal/impersonal distinction, but their critique is better understood as a critique of (one piece of evidence for) the dual-process theory. Their key finding is that there is no RT difference between utilitarian and deontological judgments in response to high-conflict “personal” dilemmas. This is a challenge for the dual-process theory regardless of whether “personal” is a good way to characterize these dilemmas. The personal/impersonal distinction is effectively irrelevant to their critique. Even if the personal/impersonal distinction had perfectly characterized the essential differences between our two sets of stimuli, identifying precisely those features of the *footbridge* and similar dilemmas that elicit disapproval, McGuire et al.’s results would still pose a challenge to the dual-process theory.

This challenge, however, has been met by a series of more recent studies, including the cognitive load study described above (Greene et al., 2008), that support the dual-process theory without depending on the personal/impersonal distinction. Greene et al. (2004) showed that utilitarian judgments, as compared to characteristically deontological judgments, are associated with increased activity in the dorsolateral prefrontal cortex (DLPFC), a brain region associated with cognitive control (Miller & Cohen, 2001). This comparison was made *within* high-conflict “personal” dilemmas (in this case defined by RT on a trial-by-trial basis) and did not involve “impersonal” dilemmas at all. Thus, while these dilemmas were labeled “personal,” the label could change without changing the implications of the result. Three studies of individual differences in cognitive style/ability also support the dual-process theory, associating utilitarian judgments with greater “need for cognition” (Bartels, 2008), “cognitive reflection” (Hardman, 2008), and working memory capacity (Moore et al., 2008). Other studies support the dual-process theory by implicating emotional responses in characteristically deontological judgments. Three neuropsychological studies (Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Koenigs et al., 2007; Mendez, Anderson, & Shapira, 2005) have found that patients with emotion-related neurological deficits make more utilitarian judgments. Along similar lines, Valdesolo and DeSteno (2006) found that inducing positive emotion elicits more utilitarian judgment. The above studies use one or more “impersonal” dilemmas as controls, but their conclusions do not

depend on the personal/impersonal distinction as drawn by Greene et al. (2001). Nor do they depend on results from low-conflict “personal” dilemmas, as in Greene et al.’s (2001) RT effect. All of these results are generated by comparisons *within* one or more high-conflict personal dilemmas. Thus, they support the dual-process theory without depending on the personal/impersonal distinction and without the item-based methodological problem identified by McGuire and colleagues.

McGuire and colleagues adduce several reasons to dismiss the evidence described above, but these arguments are scattershot and not well supported. As noted above, the study that most directly addresses their critique (Greene et al., 2008) is completely ignored, as are the published individual differences data (Bartels, 2008; Moore et al., 2008). McGuire and colleagues dismiss Greene et al.’s (2004) more recent fMRI data based on generic concerns about the cognitive interpretation of fMRI data. They raise non-specific doubts about our observed results in the anterior cingulate cortex, and make no reference at all to our interpretation of the DLPFC activity that was specifically predicted and observed in association with utilitarian judgments. McGuire and colleagues dismiss two other studies (Mendez et al., 2005; Valdesolo & DeSteno, 2006) on the grounds that they employed only the *switch* and *footbridge* dilemmas, which differ in ways other than those highlighted by the original personal/impersonal distinction. This objection reflects McGuire and colleagues’ conflation of the dual-process theory with the personal/impersonal distinction. As explained above, these two studies provide evidence for the dual-process theory that is independent of the personal/impersonal distinction. McGuire and colleagues dismiss Ciaramelli et al.’s (2007) study on the grounds that the their dilemmas might have included some of the low-conflict personal dilemmas, but they offer no explanation for why these dilemmas would generate the observed effect, which was specifically predicted by the dual-process theory. They acknowledge that the striking results observed by Koenigs and colleagues (2007), with ventromedial prefrontal patients making approximately five times more utilitarian judgments than control subjects, are not susceptible to these item-based concerns. Instead, these results are dismissed by appeal to an argument (Moll & Oliveira-Souza, 2007), mistaken, in my opinion (Greene, 2007b), to the effect that a single-system theory of moral judgment can explain why damage to the ventromedial prefrontal cortex leads to abnormal moral judgment, but leaves utilitarian moral thinking intact.

McGuire and colleagues recommend the use of more tightly controlled stimuli to better identify features of dilemmas and actions that affect people’s judgments. This a good suggestion, and one that we have implemented in more recent work aimed at replacing the personal/impersonal distinction with something better (Greene et al., *in press*), but this recommendation is orthogonal to their critique of the dual-process theory. We need not know how, exactly, the *footbridge* and *switch* dilemmas differ in order to know that they engage dissociable processing systems. McGuire and colleagues recommend the use of item analyses. I concur, and note that at least two recent studies show effects predicted by the dual-process theory consistently across items (Greene et al., 2008; Koenigs et al., 2007).

While there is much convergent evidence to support the dual-process theory, McGuire and colleagues’ critique leaves a lingering question: If the dual-process theory is correct, why do utilitarian judgments not take longer? Recent results offer a clue. In a follow-up analysis of our cognitive load data (Greene et al., 2008), we divided participants into two groups (“high-utilitarian” and “low-utilitarian”) based on their frequencies of utilitarian judgments. Both groups exhibited the critical interaction between load and utilitarian judgment, as predicted by the

dual-process theory. However, among the high-utilitarian subjects, utilitarian judgments were faster than non-utilitarian judgments in the absence of load, while the opposite was true of low-utilitarian subjects. Thus, the low-utilitarian subjects, but not the high-utilitarian subjects, exhibited a genuine RT effect of the kind reported by Greene et al. (2001). Moreover, in these more recent data we found a robust negative correlation between a participant’s tendency toward utilitarian judgment and that participant’s mean RT for utilitarian judgments in the absence of load. We found no such correlation for non-utilitarian judgments and judgments under load. This suggests that there is an additional process that drives down RT in utilitarian subjects in the absence of load. If this is correct, then an expanded version of the dual-process theory incorporating individual differences may be able to account for McGuire et al.’s results. We leave this as a matter for future research.

In sum, McGuire and colleagues have made an important contribution to research in moral psychology by definitively identifying a flaw in the RT data my colleagues and I presented in our first fMRI study. We presented these data as supporting our dual-process theory, but McGuire and colleagues have shown that they provide no such support. That said, McGuire and colleagues conflate the dual-process theory of moral judgment with the personal/impersonal distinction, too hastily dismiss more recent convergent evidence for the dual-process theory, and completely ignore the evidence that bears most directly on the issues they raise. Despite these disagreements, I admire the perspicacity with which McGuire and colleagues have conducted their analysis. Moreover, I appreciate the opportunity they have given me to address these issues and have no doubt that their efforts will advance our field.

References

- Bartels, D. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*, 381–417.
- Fischer, J. M., & Ravizza, M. (Eds.). (1992). *Ethics: Problems and principles*. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.
- Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, *2*(2), 84–92.
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*(12), 1082–1089.
- Greene, J. D. (2007a). The secret joke of Kant’s soul. In W. Sinnott-Armstrong (Ed.), *Moral psychology. The neuroscience of morality emotion disease and development* (Vol. 3). Cambridge, MA: MIT Press.
- Greene, J. D. (2007b). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323 (author reply 323–324).
- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (in press). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*.
- Greene, J., Morelli, S., Lowenberg, K., Nystrom, L., & Cohen, J. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–1154.
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400.
- Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108.
- Hardman, D. (2008). Moral dilemmas: Who makes utilitarian choices? Unpublished manuscript.
- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908–911.
- Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cognitive and Behavioral Neurology*, *18*(4), 193–197.
- Mikhail, J. (2000). Rawls’ linguistic analogy: A study of the generative grammar model of moral theory described by John Rawls in a theory of justice. Unpublished doctoral dissertation, Cornell University.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.
- Moll, J., & Oliveira-Souza (2007). Moral judgments, emotions, and the utilitarian brain. *Trends in Cognitive Sciences*.

- Moore, A., Clark, B., & Kane, M. (2008). Who shalt not kill?: Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*(6), 549–557.
- McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology*, doi:10.1016/j.jesp.2009.01.002.
- Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology*, *64*, 467–478.
- Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, *15*, 165–184.
- Schaich Borg, J., Hynes, C., Van Horn, J., Grafton, S., & Sinnott-Armstrong, W. (2006). Consequences, action, and intention as factors in moral judgments: An fMRI investigation. *Journal of Cognitive Neuroscience*, *18*(5), 803–817.
- Thomson, J. (1985). The trolley problem. *Yale Law Journal*, *94*(6), 1395–1415.
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*(6), 476–477.
- Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, *18*(3), 247–253.