Social Cognitive and Affective Neuroscience Advance Access published August 23, 2013

doi:10.1093/scan/nst102 SCAN (2013) 1 of 4

Are 'counter-intuitive' deontological judgments really counter-intuitive? An empirical reply to Kahane *et al.* (2012)

Joseph M. Paxton,[1] Tommaso Bruni,[2] and Joshua D. Greene[1]

[1]Department of Psychology, Harvard University, Cambridge, MA 02138, USA and [2]European Institute of Oncology, Via Adamello 16, 20139 Milano, Italy.

A substantial body of evidence indicates that utilitarian judgments (favoring the greater good) made in response to difficult moral dilemmas are preferentially supported by controlled, reflective processes, whereas deontological judgments (favoring rights/duties) in such cases are preferentially supported by automatic, intuitive processes. A recent neuroimaging study by Kahane *et al.* challenges this claim, using a new set of moral dilemmas that allegedly reverse the previously observed association. We report on a study in which we both induced and measured reflective responding to one of Greene *et al.*'s original dilemmas and one of Kahane *et al.*'s new dilemmas. For the original dilemma, induced reflection led to more utilitarian responding, replicating previous findings using the same methods. There was no overall effect of induced reflection for the new dilemma. However, for both dilemmas, the degree to which an individual engaged in prior reflection predicted the subsequent degree of utilitarian responding, with more reflective subjects providing more utilitarian judgments. These results cast doubt on Kahane *et al.*'s conclusions and buttress the original claim linking controlled, reflective processes to utilitarian judgment and automatic, intuitive processes to deontological judgment. Importantly, these results also speak to the generality of the underlying theory, indicating that what holds for cases involving utilitarian physical harms also holds for cases involving utilitarian lies.

Keywords: automatic processing; controlled processing; intuition; moral judgment; reflection

INTRODUCTION

Researchers have examined the cognitive and neural bases of moral judgment using neuroimaging (Greene *et al.*, 2001, 2004), lesion methods (Mendez *et al.*, 2005; Ciaramelli *et al.*, 2007; Koenigs *et al.*, 2007), psychopharmacology (Crockett *et al.*, 2010) and more traditional behavioral methods (Valdesolo and DeSteno, 2006; Greene *et al.*, 2008, 2009; Paxton *et al.*, 2011; Amit and Greene, 2012). These studies reveal a consistent pattern (Greene, 2007, 2009, 2013): controlled processes preferentially support utilitarian judgments (favoring the greater good over conflicting rights/duties), whereas automatic processes preferentially support deontological judgments (favoring rights/duties over the greater good). In other words, when utilitarian and deontological considerations conflict, deontological judgments ('It's wrong to kill one to save five',) are relatively intuitive while utilitarian judgments ('It's better to save more lives',) are relatively counter-intuitive.

In an article recently published in Social Cognitive and Affective Neuroscience, Kahane *et al.* (2012) argue that this empirical generalization does not hold. They claim that the hypothetical moral dilemmas developed by Greene *et al.* (2001) and used in subsequent studies by others, confound two independent factors: (i) the content of the moral judgment (deontological *vs* utilitarian) and (ii) the intuitiveness (*vs* counter-intuitiveness) of the judgment (c.f., Kahane, 2012). In other words, what appears to be a psychological relationship—deontological judgments are more intuitive, utilitarian judgments are more counter-intuitive—is actually an artifact produced by a narrow selection of testing materials.

To test their hypothesis, Kahane *et al.* created a new set of hypothetical moral dilemmas that were designed to reverse the previously observed content/process association, pitting an intuitive utilitarian alternative against a counter-intuitive deontological alternative. For example, these dilemmas include a case of a 'white lie' that promotes the greater good (utilitarian), but that violates the (deontological) prohibition against lying. According to Kahane *et al.*, favoring the utilitarian 'white lie' is the more intuitive response, which is not implausible. [Indeed, Greene (2007) speculated that 'white lie' cases, such as one famously described by Kant (1797/1966), might be exceptions to the previously observed pattern.]

Kahane *et al.* claim to have reversed the previously observed pattern using their new dilemmas. They present behavioral and functional magnetic resonance imaging (fMRI) evidence in support of this claim. In this brief article we make two points: first, the evidence presented by Kahane *et al.* is mixed at best. Second, we present contrary evidence using one of Kahane *et al.*'s 'white lie' dilemmas, showing that the pattern does not reverse and that it instead conforms to the opposite, previously observed pattern.

First, we consider Kahane *et al.*'s behavioral evidence. They presented their new dilemmas and several of Greene *et al.*'s original dilemmas to 18 subjects, asking them to provide an 'immediate, unreflective' response to each dilemma. These 'unreflective' responses were taken to be the product of automatic, intuitive processes. A response to a particular dilemma was thus classified as 'intuitive' when a large majority of subjects (at least 12 out of 18) gave that response to that dilemma. This resulted in a set of 'Utilitarian Intuitive' (UI) dilemmas (primarily, Kahane *et al.*'s new dilemmas) and a set of 'Deontological Intuitive' (DI) dilemmas (primarily Greene *et al.*'s original dilemmas). This classification method is likely to be unreliable. Asking subjects to give 'immediate, unreflective' responses does not mean that they will do so, as people often lack introspective access to their judgment processes (Nisbett and Wilson, 1977). Consequently, high agreement would not imply that the judgment is intuitive. For example, most people approve of killing one person to save millions of lives (Paxton *et al.*, 2011), and would likely continue to do so even when instructed to provide an immediate

Received 5 October 2012; Revised 16 January 2013; Accepted 16 July 2013

This work was supported by a National Science Foundation Graduate Research Fellowship (to J.M.P.); NSF SES-082197 8 (to J.D.G.); the FIRC Institute of Molecular Oncology and the Umberto Veronesi Foundation (to T.B.).

Correspondence should be addressed to Joseph M. Paxton, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138, USA. E-mail: jpaxton@wjh.harvard.edu

response. However, the evidence suggests that this judgment is relatively counter-intuitive.

In addition, Kahane *et al.* also examined reaction times. Consistent with their expectations, deontological judgments took longer than utilitarian judgments in response to the new UI dilemmas. However, both utilitarian and deontological responses to the UI dilemmas were on average faster than both types of responses to the 'DI' dilemmas. More specifically, the new 'counter-intuitive' deontological judgments were faster than the old intuitive deontological judgments. Thus, the reaction time data, viewed more broadly, are actually equivocal with respect to the (counter-) intuitiveness of the deontological response.

Finally, Kahane *et al.* presented fMRI data. Here, the critical contrast is between deontological and utilitarian responses to the UI dilemmas. Previous research associates activity in the dorsolateral prefrontal cortex (DLPFC) with counter-intuitive moral judgment (Greene *et al*, 2004; Cushman *et al.*, 2011). Thus, if Kahane *et al.*'s hypothesis is correct, they should observe increased DLPFC activity for deontological judgments (relative to utilitarian judgments) in response to UI cases. They did not. Instead, they observed increased activity in the anterior cingulate cortex, which, though part of the prefrontal control network, is not specifically associated with the application of cognitive control (MacDonald *et al.*, 2000).

Thus, the evidence for Kahane *et al.*'s reversal featuring 'counter-intuitive' deontological judgments is mixed at best. Here we use the 'Cognitive Reflection Test' (CRT; Frederick, 2005) to induce and measure responses that are more reflective and less intuitive (Paxton *et al.*, 2011; Pinillos *et al.*, 2011). As this method employs an implicit experimental manipulation, rather than relying on subjects' ability to modify their judgment processes, or on the assumption that agreement implies intuitiveness, it offers a more diagnostic test of the nature of the underlying processes that support deontological and utilitarian moral judgments.

METHODS, STIMULI AND PARTICIPANTS

The CRT consists of three questions that elicit incorrect, intuitive responses, which can be overridden by correct, reflective responses through the application of basic math, e.g.,

> A bat and a ball cost $1.10.
> The bat costs one dollar more than the ball.
> How much does the ball cost?

People reliably have the intuition that the ball costs $0.10. However, some basic math and a bit of reflection reveal that the correct answer is $0.05. Although many subjects give the intuitive response for all three questions, more than half in Frederick's (2005) sample gave the reflective response to at least one of the questions.

In a previous study (Paxton *et al.*, 2011), we found that correctly answering at least one CRT question reinforced the value of reflection, subsequently leading subjects to make more utilitarian judgments in response to several of Greene *et al.*'s (2001) original dilemmas. In addition, we found that answering more CRT problems correctly caused judgments to be more utilitarian. (In a control condition the order of the tasks was reversed, establishing a baseline moral judgment rating.) Here we use the same method in a 2×2 design: subjects completed the three CRT items either before (CRT-First) or after (Dilemma-First) responding to a moral dilemma. The dilemma was either a standard 'Deontological Intuitive' (DI) dilemma ('Sophie's Choice' from Greene *et al.*, 2001) or one of Kahane *et al.*'s new 'Utilitarian Intuitive' (UI) dilemmas ('White Lie 2').[1]

[1]Note that both dilemmas were used by Kahane *et al.* (2012). We follow their category labels (DI and UI) for convenience rather than to indicate our agreement with the categorization.

This design provides a direct test of Kahane *et al.*'s alternative hypothesis: according to Kahane *et al.*, successful CRT performance should lead to (or be associated with) increased deontological responding in the UI case. However, if Greene *et al.*'s original dual-process theory is correct, successful CRT performance should lead to (or be associated with) increased utilitarian responding in the UI case as well as the more standard DI case. This design also gives us the opportunity to replicate previous results (Paxton *et al.*, 2011) using a previously unused standard DI case.

The full text of both dilemmas follows:

> Sophie's Choice: It is wartime and you and your two children, ages eight and five, are living in a territory that has been occupied by the enemy. At the enemy's headquarters is a doctor who performs painful experiments on humans that inevitably lead to death. He intends to perform experiments on one of your children, but he will allow you to choose which of your children will be experimented upon. You have twenty-four hours to bring one of your children to his laboratory. If you refuse to bring one of your children to his laboratory he will find them both and experiment on both of them. Should you bring one of your children to the laboratory in order to avoid having them both die?
>
> White Lie 2: A young friend of yours always greatly admired his uncle, who has just died and whom you knew well. At the funeral the nephew asks you to tell him what his uncle really thought of him. As a matter of fact, his uncle disliked him and the young man would be devastated to find this out. However, his uncle was superficial and spiteful in his opinions of people and was not worthy of the young man's esteem. It would do much good for the young man's confidence and self esteem if he thought that his uncle thought well of him. Should you tell your friend that his uncle disliked him?

Subjects responded using a Likert scale that ranged from 1 (Definitely Shouldn't) to 7 (Definitely Should). Thus, in the DI scenario, a lower rating corresponds to the deontological judgment that you should not sacrifice one of your children to save the other, even if this means that they will both die, while a higher rating corresponds to the utilitarian judgment that you should sacrifice one to save the other. In the UI scenario, a lower rating corresponds to the utilitarian judgment that you should lie to your friend to protect his feelings, while a higher rating corresponds to the deontological judgment that you should tell your friend the truth, even if doing so would devastate him.

Subjects (44 females, 17 males, 4 gender unspecified; mean age = 35.43, s.d. = 14.12, 5 age unspecified) were recruited from Craigslist.com, and participated on a voluntary basis in a brief web-based study.

RESULTS

We replicated both of the main findings from Paxton *et al.* (2011) using the DI dilemma (Sophie's Choice). Completing the CRT first led to more utilitarian judgments, for subjects who answered at least one CRT question correctly (CRT-First $M = 5.5$, Dilemmas-First $M = 3.0$, $t(15) = 2.2$, $p = .04$). The effect fell short of significance when including all subjects (CRT-First $M = 3.67$, Dilemmas-First $M = 2.48$, $t(36) = 1.68$, $P = 0.1$). This is consistent with previous results and was expected as this latter analysis includes subjects who showed no evidence of having reflected on the CRT questions. Importantly, the proportion of subjects in each condition did not differ statistically before and after exclusion [Pre-Exclusion CRT-First: 15 of 38 (39%), Post-Exclusion CRT-First: 6 of 17 (35%),

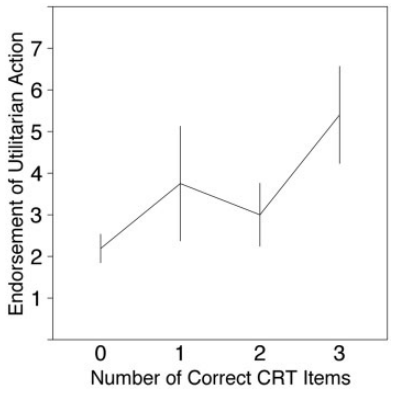

Fig. 1 Mean endorsement of the utilitarian action ('sacrifice one child to avoid the deaths of both children') in the 'Deontological Intuitive' dilemma (Sophie's Choice), broken down by the number of CRT items answered correctly ($r = 0.44$, $P = 0.006$). Error bars represent standard error of the mean.

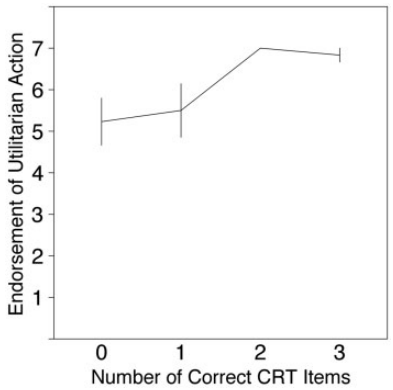

Fig. 2 Mean endorsement of the utilitarian action ('lie to your friend to protect his feelings') in the 'Utilitarian Intuitive' dilemma (White Lie 2), broken down by the number of CRT items answered correctly ($r = 0.45$, $P = 0.02$). Error bars represent standard error of the mean.

$\chi^2 = 0$, $P = 1$]. In addition, subjects made more utilitarian judgments the more CRT questions they answered correctly ($r = 0.44$, $P = 0.006$; Figure 1).[2] As in the original study, this effect was driven mainly by subjects in the CRT-First condition ($r = 0.57$, $P = 0.03$; Dilemmas-First, $r = 0.24$, $P = 0.27$), suggesting that the correlation was at least partially induced by the CRT manipulation.

For the UI dilemma (White Lie 2), we reverse coded the Likert scale ratings to make them consistent with those of the DI dilemma (higher = more utilitarian). The effect of condition was non-significant, both for subjects who answered at least one CRT question correctly (CRT-First $M = 6.63$, Dilemmas-First $M = 6.33$, $t(12) = 0.56$, $P = 0.59$), and when including all subjects (CRT-First $M = 5.77$, Dilemmas-First $M = 6.0$, $t(25) = -0.35$, $P = 0.73$). Again, the proportion of subjects in each condition did not change significantly after exclusion [Pre-Exclusion CRT-First: 13 of 27 (48%), Post-Exclusion CRT-First: 8 of 14 (57%), $\chi^2 = 0.05$, $P = 0.83$]. However, the more sensitive correlational test revealed that subjects were more utilitarian, the more CRT questions they answered correctly ($r = 0.45$, $P = 0.02$; Figure 2).[3] Once more, this effect was driven primarily by subjects in the CRT-First condition ($r = 0.61$, $P = 0.03$; Dilemmas-First $r = 0.3$, $P = 0.29$), suggesting a partially induced correlation.

Finally, because both CRT performance and moral judgment may be sensitive to demographic variables such as age and sex (e.g., Frederick, 2005), we repeated the correlational analyses above, controlling for both of these variables. The positive association between CRT performance and utilitarian judgment survived controls for age and sex, together and separately, in both the UI and DI dilemmas (all $P < 0.05$).

DISCUSSION

Studies of moral cognition using a wide range of methods reveal a consistent pattern: when individual rights/duties conflict with the greater good, deontological judgments favoring rights/duties are more intuitive (more automatic), while utilitarian judgments favoring the greater good are more counter-intuitive (more controlled). Kahane *et al.* (2012) claim to have constructed cases that reverse this pattern, UI dilemmas in which the utilitarian response is more intuitive and the deontological response is more counter-intuitive. We have raised doubts about the behavioral and fMRI evidence presented in support of this claim. More importantly, we have provided positive evidence against it.

Building on previous work (Pinillos *et al.*, 2011; Paxton *et al.*, 2011), we used the CRT (Frederick, 2005) to induce and measure moral judgments that are more reflective and less intuitive. We examined more closely one of Kahane *et al.*'s UI dilemmas, which was explicitly designed to reverse the standard pattern. Not only did the standard pattern fail to reverse, it was observed where it was least expected to be found (Greene, 2007): more reflective individuals were more willing to approve of a utilitarian white lie, just as they are more willing to approve of a utilitarian physical harm (Paxton *et al.*, 2011). The effect of induced reflection on the UI dilemma was non-significant, but this may be due to a ceiling effect. Judgments in response to this dilemma were highly utilitarian at baseline leaving little room for increase. Nevertheless, the more sensitive correlational test revealed a significant positive relationship between reflective CRT responding and utilitarian judgment, consistent with Greene *et al.* (2001, 2004, 2008) and directly opposed to the alternative theory advanced by Kahane *et al.* (2012). Thus, the present results provide the strongest evidence to date for the generality of Greene *et al.*'s dual-process theory of moral judgment.

One might ask whether these effects are due to affect-related confounds. For example, a previous study has found that inducing positive affect increases utilitarian responding, presumably by counteracting the negative affective responses that are hypothesized to drive deontological moral judgments (Valdesolo and DeSteno, 2006). We addressed this alternative explanation in previous work (Paxton *et al.*, 2011) and found that the CRT fails to induce either positive or negative affect.

In addition, one might wonder whether the relationship between CRT performance and utilitarian judgment can be explained simply by appeal to the numerical content of both types of questions. That is, responding correctly to CRT questions requires one to perform basic calculations, just as utilitarian moral responding does (Kahane, 2012). Although the present results do not rule out this explanation in the case of the DI dilemma (*Sophie's Choice*), which requires an explicit numerical calculation, the appeal to numerical cognition cannot explain the results in the case of the UI dilemma (*White Lie 2*), which includes no numerical content. This point is crucial, as it is the characterization of the 'UI' dilemma that is in question.

Finally, we note that we have presented evidence concerning only one of Kahane *et al.*'s UI dilemmas. Thus, it is possible that one or more of these dilemmas reverses the pattern widely observed in

[2]Notably, more than twice as many subjects scored 0 on the CRT ($n = 21$) than scored 1 ($n = 4$), 2 ($n = 8$), or 3 ($n = 5$). We ran an additional analysis employing a binary coding for CRT scores (less than *vs* greater than 0), allowing us to compare two comparably sized groups of 'unreflective' (0) and 'reflective' (1) subjects. The correlation between utilitarian judgment and reflectiveness remained significant ($r = 0.39$, $P = 0.02$).

[3]Again, more than twice as many subjects scored 0 on the CRT ($n = 13$) than scored 1 ($n = 4$), 2 ($n = 4$) or 3 ($n = 6$). As before, we recoded all those scoring 1–3 as 1, yielding a binary CRT score. Once again, the correlation between utilitarian judgment and reflectiveness remained significant ($r = 0.39$, $P = 0.046$).

previous studies of moral cognition. Nevertheless, there is currently no compelling evidence for such a reversal. Instead, the evidence increasingly supports our claim that controlled, reflective processes preferentially support utilitarian judgments, whereas automatic, intuitive processes preferentially support deontological judgments.

Conflict of Interest

None declared.

REFERENCES

Amit, E., Greene, J.D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science*, *23*, 861–8.

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.

Crockett, M., Clark, L., Hauser, M., Robbins, T. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences,* USA, *107*(40), 17433–8.

Cushman, F., Murray, D., Gordon-McKeon, S., Wharton, S., Greene, J.D. (2011). Judgment before principle: engagement of the frontoparietal control network in condemning harms of omission. *Social Cognitive and Affective Neuroscience*, *7*(8), 888–95.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*(4), 25–42.

Greene, J.D. (2007). The secret joke of Kant's soul. In: Sinnott-Armstrong, W., editor. *Moral Psychology*, Vol. 3. Cambridge, MA: MIT Press.

Greene, J.D. (2009). The cognitive neuroscience of moral judgment. In: Gazzaniga, M.S., editor. *The Cognitive Neurosciences* 4th edn. Cambridge, MA: MIT Press, pp. 987–99.

Greene, J.D. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. New York: Penguin Press.

Greene, J.D., Cushman, F.A., Stewart, L.E., Lowenberg, K., Nystrom, L.E., Cohen, J.D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364–71.

Greene, J.D., Morelli, S.A., Lowenberg, K., Nystrom, L.E., Cohen, J.D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition*, *107*(3), 1144–54.

Greene, J.D., Nystrom, L., Engell, A., Darley, J., Cohen, J. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400.

Greene, J.D., Sommerville, R., Nystrom, L., Darley, J., Cohen, J. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–8.

Kant, I. (1797/1966). On a supposed right to lie from philanthropy. In: Gregor, M.J., editor. *Practical Philosophy*. Cambridge: Cambridge University Press, pp. 611–5.

Kahane, G. (2012). On the wrong track: Process and content in moral judgment. *Mind and Language*, *27*, 519–545.

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgments. *Social Cognitive and Affective Neuroscience*, *7*(4), 393–402.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*(7138), 908–11.

MacDonald, A.W., Cohen, J.D., Stenger, V.A., Carter, C.S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*, 1835–8.

Mendez, M., Anderson, E., Shapira, J. (2005). An investigation of moral judgement in frontotemporal dementia. *Cognitive and Behavioral Neurology*, *18*(4), 193–7.

Nisbett, R.E., Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–59.

Paxton, J.M., Ungar, L., Greene, J.D. (2011). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*, 163–77.

Pinillos, N., Smith, N., Nair, G., Marchetto, P., Mun, C. (2011). Philosophy's new challenge: Experiments and intentional action. *Mind and Language*, *26*(1), 115–39.

Valdesolo, P., DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*(6), 476–7.