The Effect of Probability Anchors on Moral Decision Making

Chris M. Brand (c.brand@bbk.ac.uk)
Department of Psychological Sciences, Birkbeck, University of London
Malet Street, London, WC1E 7HX, UK

Mike Oaksford (m.oaksford@bbk.ac.uk)
Department of Psychological Sciences, Birkbeck, University of London
Malet Street, London, WC1E 7HX, UK

Abstract

The role of probabilistic reasoning in moral decision making has seen relatively little research, despite having potentially profound consequences for our models of moral cognition. To rectify this, two experiments were undertaken in which participants were presented with moral dilemmas with additional information designed to anchor judgements about how likely the dilemma's outcomes were. It was found that these anchoring values significantly altered how permissible the dilemmas were found when they were presented both explicitly and implicitly. This was the case even for dilemmas typically seen as eliciting deontological judgements. Implications of this finding for cognitive models of moral decision making are discussed.

Keywords: cognitive science; decision making; experimental research with adult humans; moral decision making; psychology; reasoning; social cognition

Introduction

Scientific interest in morality has increased dramatically over the past decade, with the rationalist cognitive-developmental theories (Kohlberg, 1981) that have historically dominated moral psychology declining in popularity compared to theories which propose a greater role for intuition and affect (Greene, Sommerville, Nystrom, Darley & Cohen, 2001; Haidt, 2001). Despite this heightened level of activity within the field, the science of morality is not without its problems. Demand for theories formulated at a computational level of analysis (Marr, 1982) has intensified in recent years (Mikhail, 2011), as well as for research which more directly connects moral psychology to related subjects such as decision science and the psychology of reasoning. Fiddick (2004), for instance, notes that the cognitive and moral reasoning literatures were so divorced from one another that they had reached two opposing consensuses about whether deontic reasoning is a unified phenomenon or not.

One of the most influential recent developments within the psychology of reasoning has been the claim that people do not reason as if the premises of an argument were certain; instead, they reason about them in a probabilistic fashion. Despite this "new paradigm" (Over, 2009) having potentially major implications for moral psychology, it has received little explicit attention within the field. Indeed, there is experimental evidence from the moral psychological literature suggesting that probabilistic inference may play an important role in moral judgement (Greene, Cushman, Stewart, Lowenberg, Nystrom & Cohen, 2009; Liu & Ditto, 2013), although this possibility is only rarely treated as theoretically important (Cushman, 2013; Sloman et al., 2009). Explicit investigation of whether individuals do not take the stated premises of moral dilemmas for granted would certainly be productive; if it turns out to be the case that they do not, then any models of moral decision making that we produce should take this into account.

If individuals do indeed reason about moral dilemmas in a manner which incorporates background assumptions of probability, however, then studying this may prove difficult without attempting to homogenize these assumptions between individuals in some manner. Although participants could be asked directly how likely they find certain outcomes in a given moral dilemma, these probability estimates may vary drastically between individuals. The introduction of anchoring values may, however, provide one possible method of compensating for this. The effect of anchors on judgement and decision making is a well-studied topic (Kahneman, 1992); reference points can significantly alter a numerical judgement, even when that reference point is in a domain entirely unrelated to that of the numerical judgement being elicited (Tversky & Kahneman, 1974; Englich, Mussweiler and Strack, 2006). Previous research has found, most relevantly, that anchors can also have a large effect on the subjective probabilities associated with an event (Wright & Anderson, 1989); based upon this, it seems plausible that an anchoring value may be used to set a baseline for probabilistic reasoning to take place.

To investigate the role that probabilistic reasoning plays in moral decision making, an experiment was conducted which sought to investigate what effect the presence of anchoring values has on permissibility judgements regarding artificial moral dilemmas. It was predicted that there would be significant effects of dilemma and probability on participants' permissibility judgements. It was also predicted that there would be a significant interaction between dilemma and probability.

Experiment 1

Method

Design A repeated measures experimental design was employed. There were two independent variables employed; the moral dilemma described, with 5 levels (trolley,

footbridge, fumes, transplant and control), and the probability level of the dilemma, with 4 levels (low, medium, high, and certain). The dependent variable was judged permissibility, measured on a 7 point scale where one extreme was labelled as “forbidden” and the other as “obligatory”.

Participants 260 participants were recruited from online sources. 35 participants failed to fully complete the experiment and were thus excluded from eventual analysis, for a total of 225 participants. 54.7% reported their gender as female, with 3.6% reporting their gender as "other". 51.6% reported that they had studied psychology in an academic setting, and 30.2% that they had studied moral philosophy. Participation was entirely voluntary.

Procedure Participants completed the study online. On the first page of the website hosting the study, the participants were presented with an introductory screen outlining the broad aim and methodology of the study. Participants were informed that any answers that they provided would be completely anonymous and that they were free to withdraw from participation at any point, and were then asked for their consent to take part in the experiment. Once the participants had given this, they were taken to the first of twenty dilemmas.

In each dilemma, participants were presented with one of five moral dilemmas based upon a selection of those found in Greene et al. (2001); the trolley problem, the footbridge problem, the “fumes” problem (in which a hospital attendant must choose whether to divert poisonous fumes from a room containing five patients to a room containing one patient), the transplant problem (in which a doctor must choose whether to kill a healthy individual in order to transplant their organs into five dying patients), and a costless control problem (in which a doctor must choose whether to administer medicine that may be past its use-by data). In each instance of a dilemma, the participants were also given two further pieces of information. The first was an “anchor”; in the footbridge problem, for instance, they were informed that a weight of 30 stone has a 50% chance of stopping a train successfully, while in the fumes problem they were informed that ten minutes after the accident that released the fumes there will be a 50% chance that a lethal level of fumes will have already entered the room containing the five patients. The second piece of information given to participants was intended to either make the planned outcome of the dilemma’s action seem *less* probable than the anchor (e.g., in the fumes problem, it had been fifteen minutes since the accident), *equally* probable (it had been ten minutes since the accident), or *more* probable (it had been five minutes). Additionally, there was a fourth “certain” condition where information about the anchor was removed entirely and the participants were only presented with the core text of the dilemma. It was presumed that this would be seen as the most probable condition, being equivalent to the dilemma’s outcomes occurring as stated with a probability of 1.

Below the presentation of the dilemma, the participants were asked to rate the permissibility of the dilemma’s proposed action on a seven point scale, with 1 labelled as “forbidden” and 7 as “obligatory”. The participants were each presented with the four variants of the five moral dilemmas in a randomized order, and once these twenty trials had all been completed then they were asked for their demographic data and finally thanked for their time.

Results

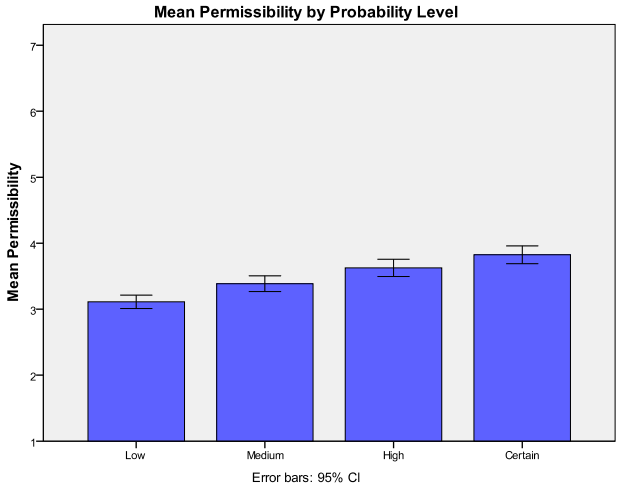

Figure 1: The mean permissibility of each probability level averaged across all dilemmas in experiment 1. Error bars are 95% confidence intervals.

A 5 (dilemma) x 4 (probability level) repeated measures ANOVA was performed on the data in order to assess the effect that both varying the probability and dilemma had on participants’ permissibility judgements. Mauchly’s test indicated that the assumption of sphericity had been violated by the dilemma variable, probability variable and interaction between them; the Greenhouse-Geisser correction was used to accommodate for this. There was a significant main effect of dilemma ($F(4,896) = 798.72$, $p < 0.001$, $\eta_p^2 = 0.78$), a significant main effect of probability ($F(3,672) = 151.44$, $p < 0.001$, $\eta_p^2 = 0.40$) and a significant interaction between dilemma and probability ($F(12,2688) = 32.33$, $p < 0.001$, $\eta_p^2 = 0.13$). Post-hoc Bonferroni-corrected comparisons revealed that the judged permissibility of all levels of probability differed significantly, as did the judged permissibility of all dilemmas employed.

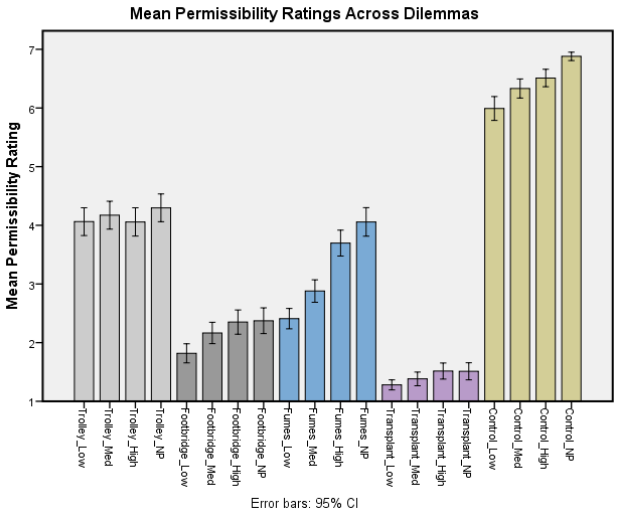

Figure 2: The mean permissibility of each probability level (low, medium, high and non-probabilistic) across all dilemmas (trolley, footbridge, fumes, transplant and control) in experiment 1. Error bars are 95% confidence intervals.

Discussion

The results from the study supported the hypotheses; there was a significant effect of probability on permissibility judgements, a significant effect of dilemma on permissibility, and there was also a significant interaction between moral dilemma and probability.

One potentially interesting finding is that the trolley problem seemed unique in being the only dilemma which was unaffected by varying probability. Quite why this might be the case is difficult to determine; in the psychological literature the trolley problem is the archetypal example of a moral dilemma which elicits naively utilitarian responses, so it is surprising that - of all the dilemmas presented in this experiment - judgements of the trolley problem seemed *least* affected by changing the implicit probability of the action's outcome. It may be that participants were unconvinced by the chosen method of attempting to vary probability; varying the length of a wire, no matter how old and frayed, may be seen as unlikely to affect the probability of a signal being sent successfully along it. This would agree with previous research which has found that anchoring values that are considered to be implausible will affect decision making to a lesser extent than more sensible anchors (Mussweiler & Strack, 2000). If we accept this explanation for the trolley problem's data, then it appears that the participants were engaging in sophisticated reasoning about whether the experiment's attempts to implicitly vary probability were plausible; at the very least, it suggests that they were not falling for the "good participant" effect (Nichols & Maner, 2008) and merely varying their judgements about the dilemmas because they believed that the experimenter wished them to do so.

It is also worth pointing out that the effect of varying probability on the transplant problem was relatively small; in this case, it is likely due to a floor effect caused by the overall impermissibility of that particular moral dilemma. If it is already seen by most participants as forbidden to kill a healthy patient so that their organs can definitely save the lives of five others, reducing the likelihood that those five patients' lives will be saved cannot make the action *more* forbidden. In the case of the control problem – where it may be argued that there should have been a similar ceiling effect – it is plausible that there was uncertainty amongst participants regarding potential negative effects of out-of-date medicine.

In order to test whether participants were indeed unconvinced by the probabilistic elements of the dilemmas, a second experiment was undertaken which sought to more strictly control participants' assumptions about the likelihood of outcomes by assigning explicit probabilities to the possible additions in each dilemma. As in the first experiment, it was predicted that there would be significant effects of dilemma and probability on participants' permissibility judgements, as well as there being a significant interaction between dilemma and probability.

Experiment 2

Method

Design A repeated measures experimental design was employed. As in the first experiment, there were two independent variables employed; the moral dilemma described, with 5 levels (trolley, footbridge, fumes, transplant and control), and the probability level of the dilemma, with 4 levels (low, medium, high, and certain). The dependent variable was again judged permissibility, measured on a 7 point scale where 1 was labelled as "forbidden" and 7 as "obligatory".

Participants 80 participants were recruited from Birkbeck's internal subject pool. 10 participants failed to fully complete the experiment and were excluded from the final analysis, leaving 70 participants to be analysed. 65.7% of participants reported their gender as female, 74.3% that they had studied psychology, and 12.9% had studied moral philosophy. Participants received course credit for their time.

Procedure Participants completed the study online. As in the first experiment, upon opening the website which hosted the study they were first exposed to an introductory screen outlining the general purpose of the study. Participants were assured of their anonymity and asked to provide consent to take part. Once this had been given, they were taken to the first of twenty dilemmas.

In each dilemma, participants were again presented with one of the five moral dilemmas used in the previous study. For each dilemma, the participants were given two further pieces of information. The first was, as in the previous study, an anchoring value. The second piece of information given was again the same as the previous study, but participants were additionally informed *explicitly* about the probability that the action carried given the extra

information. In the fumes dilemma's low probability condition, as an example, participants were informed that it had been fifteen minutes since the accident and diverting the fumes only had a 25% chance of successfully saving the five patients; in the first study, the participants would only have been told that it had been fifteen minutes since the accident. As in the previous study, there was also a "certain" condition in which the information about the anchor and the probability of the action being successful was omitted. Once participants had completed all twenty possible dilemmas in a randomized order, they were asked for their demographic data and finally thanked for their time.

Results

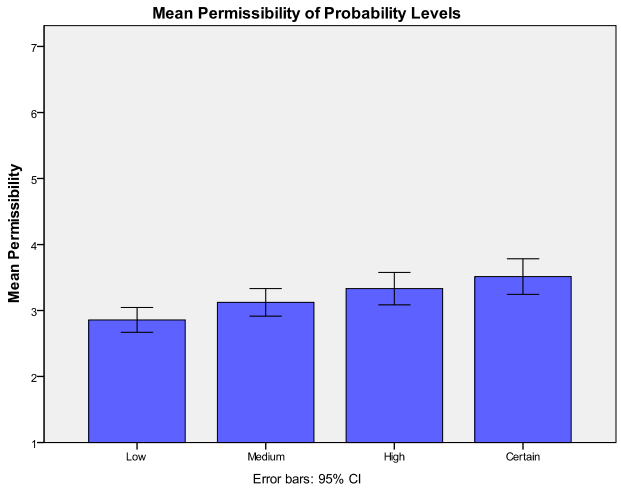

Figure 3: The mean permissibility of each probability level averaged across all dilemmas in experiment 2. Error bars are 95% confidence intervals.

A 5x4 repeated measures ANOVA was performed on the data in order to assess the effect that both varying the probability and dilemma had on participants' permissibility judgements. The Greenhouse-Geisser correction was again employed, as Mauchly's test indicated that the assumption of sphericity had been violated by both the dilemma and probability variables alongside their interaction. It was found that there was a significant main effect of dilemma ($F(4,276) = 210.27$, $p < 0.001$, $\eta_p^2 = 0.75$), a significant main effect of probability ($F(3,207) = 20.10$, $p < 0.001$, $\eta_p^2 = 0.23$) and a significant interaction between dilemma and probability ($F(12,828) = 2.494$, $p < 0.005$, $\eta_p^2 = 0.35$). Post-hoc Bonferroni-corrected comparisons revealed that the judged permissibility of all levels of probability differed significantly from each other. The trolley problem's permissibility differed significantly from all other dilemmas, while the control problem differed significantly from all other dilemmas except the fumes problem; no other dilemmas differed significantly.

Discussion

The hypotheses were again supported; there was a significant effect of probability on permissibility judgements, and there was a significant interaction between moral dilemma and probability. This result serves, to an extent, as a conceptual extension of past research which has investigated the effect of explicitly varying probability within moral dilemmas (Shenhav and Greene, 2010), but directly comparing the results of this study to the first experiment is especially informative. One notable difference is that, in this second experiment, the permissibility judgements for the various forms of the trolley problem did significantly differ. This suggests that, in the previous study, participants may indeed have been reasoning about whether the proposed anchor actually seemed plausible as a manner of altering the outcome's probabilities; it was only when explicitly informed of the new probability that this difference was observable. Once again, a floor effect was observed within the transplant problem. This raises the possibility that dilemmas which elicit such extreme reactions from participants may not be suitable for investigating variables which cause relatively subtle shifts in moral judgement.

General Discussion

Based upon the presented experiments, it appears that participants do take anchors and probability into account when engaging in moral decision making. In the first study, it was found that changing the implicit probability of a dilemma did indeed have a significant effect on permissibility judgements. The second study sought to clarify these findings; it was found that varying probability explicitly affected permissibility judgements in a more consistent fashion, suggesting that when only varying probability implicitly the effect on participants' judgements may be subtle - if not entirely unobservable, depending on the dilemma involved and how probability has been altered. Since in both experiments it was found that dilemmas typically classified as eliciting deontological responses were nonetheless influenced by varying probability, this may impact how we wish to model moral decision making; whatever processes that are responsible for deontological judgements evidently take probability into account.

There are further possible studies which may shed more light on this issue; how varying the anchoring figures themselves will effect reasoning has not been investigated, for example. It is certainly plausible, given the presented findings, that informing participants that an anchor has a higher probability of causing an outcome will lead to generally higher permissibility judgements. In particular, more investigation may be useful in order to determine exactly why there was a different effect of probability level on the permissibility of the trolley problem in each experiment; it seems possible that this was due to the participants being sensitive to the believability of the dilemma's proposed causal structure. Future research will

hopefully serve to more fully illuminate the role that probabilistic reasoning plays in moral decision making.

Acknowledgments

This research was not supported by funding from any public, commercial or not-for-profit agency.

References

Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, *17*(3), 273-292.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, *32*(2), 188-200.

Fiddick, L. (2004). Domains of deontic reasoning: resolving the discrepancy between the cognitive and moral reasoning literatures. *Quarterly Journal of Experimental Psychology*, *57*(3), 447-474.

Greene, J.D., Cushman, F.A., Stewart. L.E., Lowenberg, K., Nystrom, L.E., and Cohen, J.D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition*, *111*(3), 364-371.

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.

Haidt, J. (2001) The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–34.

Kahneman, D. (1992). Reference points, anchors, norms, and mixed feelings. Organizational behavior and human decision processes, *51*(2), 296-312.

Kohlberg, L. (1981). *Essays on moral development, volume 1: The philosophy of moral development*. New York: Harper Row.

Liu, B.S. & Ditto, P.H. (2013). What dilemmas? Moral evaluation shapes factual belief. *Social Psychological and Personality Science*, *4*(3), 316-323.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. WH San Francisco: Freeman and Company.

Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment.* New York: Cambridge University Press.

Mussweiler, T., & Strack, F. (2000). Numeric judgments under uncertainty: The role of knowledge in anchoring. *Journal of Experimental Social Psychology*, *36*(5), 495-518.

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *The Journal of general psychology*, *135*(2), 151-166.

Over, D.E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning*, *15*(4), 431-438.

Shenhav, A., & Greene, J. D. (2010). Moral judgments recruit domain-general valuation mechanisms to integrate representations of probability and magnitude. *Neuron*, *67*(4), 667-677.

Sloman, S.A., Fernbach, P.M., & Ewing, S. (2009). Causal models: the representational infrastructure for moral judgment. In D. Bartels, C. W. Bauman, L.J. Skitka, & D. Medin (Eds.), *Moral judgment and decision making: The psychology of learning and motivation* (Vol. 50). San Diego, CA: Elsevier.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Wright, W. F., & Anderson, U. (1989). Effects of situation familiarity and financial incentives on use of the anchoring and adjustment heuristic for probability assessment. *Organizational Behavior and Human Decision Processes*, *44*(1), 68-82.