Judging the morality of utilitarian actions: how poor utilitarian accessibility makes judges irrational

Petko Kusev - Kingston University London, City University London
Paul van Schaik - Teesside University
Shrooq Alzahrani - City University London
Samantha Lonigro – Kingston University London, Sapienza University of Rome
Harry Purser – University of Nottingham
Brief Report

Author Note

Petko Kusev and Paul van Schaik are supported by Economic and Social Research Council Grant RES-000-22-1768. We also thank the Nuffield Foundation (SGS36177) and The British Academy (SG47881/SG091144) for supporting Petko Kusev in his research. We are grateful to Peter Barr for his helpful comments and for programming the experiment.

Correspondence concerning this article should be addressed to Petko Kusev, Department of Psychology, Kingston University London, KT1 2EE, United Kingdom, E-mail: p.kusev@kingston.ac.uk, Phone: +44 (0)208 4177339; Fax: +44(0)20 84172292; Department of Psychology, City University, London, EC1V 0HB, United Kingdom, e-mail: p.kusev@city.ac.uk; Phone: +44(0)20 70408978; Fax: +44(0)20 70408581

Abstract

Is it acceptable and moral to sacrifice a few people's lives to save many others?

Research on moral dilemmas in psychology, experimental philosophy and

neuropsychology has shown that respondents judge utilitarian personal moral

actions (footbridge dilemma) as less appropriate than equivalent utilitarian

impersonal moral actions (trolley dilemma). Accordingly, theorists (e.g., Greene et

al., 2001) have argued that judgments of appropriateness in personal moral

dilemmas are more emotionally salient and cognitively demanding (taking more time

to be rational) than impersonal moral dilemmas. Our novel findings show an effect of

psychological accessibility (driven by partial contextual information; Kahneman,

2003) on utilitarian moral behavior and response time for rational choices. Enhanced

accessibility of utilitarian outcomes through comprehensive information about moral

actions and consequences boosted utility maximization in moral choices, with

rational choices taking less time. Moreover, our result suggests that previous results

indicating emotional interference, with rational choices taking more time to make,

may have been artifacts of presenting partial information.

Keywords: utility, moral dilemmas, accessibility, judgments, rational choice

2

Is it acceptable and moral to sacrifice a few people's lives to save many others? 'It is the greatest happiness of the greatest number that is the measure of right and wrong'. With these words, the British philosopher Jeremy Bentham (1789) defined the nature of utilitarian actions: behaviors judged as morally right only by virtue of their outcome (Bentham, 1789). From the utilitarian point of view, Bentham (1789; 1948) noted that is acceptable to sacrifice a small number of people's lives to save a greater number because this results in greater utility (happiness) overall. In contrast, deontologists (e.g., Kant, 1785; 1959) have argued that it is not acceptable, because living is a fundamental right for everyone, and no one has the right to take that from anyone, regardless of any benefits that may arise from doing so. Research in psychology, experimental philosophy and neuropsychology has revealed that moral judgments of the appropriateness of life-saving actions are not strictly utilitarian, but are influenced by the type of involvement (Greene, Sommerville, Nystrom, Darley & Cohen, 2001; Greene & Haidt, 2002; Mikhail, 2007, 2009; Thomson, 1985). In particular, directly taking action ('personal action') in scenarios (one person pushing another from the bridge in order to save several others, in the 'footbridge dilemma') was judged to be less appropriate than indirectly taking action ('impersonal action') (a person 'switching a mechanism', killing one person in order to save several others, in the 'trolley dilemma').

Various theoretical attempts have been made to account for these behavioral differences in response to personal and impersonal dilemmas. Traditionally, moral-psychology theorists have focused on the role of emotional processes in moral judgments (Cushman, Young & Hauser, 2006; Greene et al., 2001; Greene & Haidt, 2002; Nakamura, 2013; Valdesolo & DeSteno, 2006). For instance, Greene and

colleagues (2001; Greene & Haidt, 2002) found that respondents spent more time judging the appropriateness of personal moral actions than of impersonal actions. This result seems puzzling and surprising from a strict utilitarian perspective, given that the two dilemma types offer identical utility.

In an attempt to provide an account of the above result in terms of the relationship between implicit and explicit cognitive processes in moral judgments, Greene and colleagues (2001) proposed a dual-process theory of moral behavior, stating that moral judgments can be driven via both (i) implicit, fast, affective and (ii) explicit, slow, controlled psychological mechanisms (Forbes & Grafman, 2010; Greene et al., 2001; Greene & Haidt, 2002; Greene, Nystrom, Engell, Darley & Cohen, 2004; Moore, Clark & Kane, 2008). In Greene's view, the affective system is likely to be activated by 'personal' moral considerations, while the cognitive system might favor utilitarian consequences and thus rational thinking. This proposal has been supported by behavioral experiments (Greene et al., 2001), testing utilitarian choices in a morally challenging situation, in which a trolley is riding a rail and - if it proceeds on its way - five people tied on the track will be killed. The participants were presented with two different opportunities: in the trolley dilemma, to hit a switch and make the trolley change its track, killing one person tied to another rail, or to do nothing and let the five people die; similarly, in the footbridge dilemma, to push a person off the bridge and onto the tracks below where his large body will stop the trolley, saving the five people tied up onto the track, or to do nothing and let the five people die. The results show that people judge as *appropriate* sacrificing one person for the sake of five in the trolley dilemma, but judge as inappropriate sacrificing one person in order to save five in the footbridge dilemma.

According to the moral dual-process model, what makes the difference between the two types of dilemma is the degree of personal affective and cognitive involvement. Consistent with this model, participants took longer to accept (i.e. judge as appropriate) personal moral actions that would maximize utilitarian outcomes (rational moral judgments) than to reject such actions as inappropriate (irrational moral judgments) (Greene et al., 2001; Greene & Haidt, 2002). Greene and colleagues (2001) argued that when participants faced personal (footbridge-like) dilemmas in which one's moral rules conflict with the outcomes, both affective and cognitive systems were recruited. The former would favor rejecting the actions, for the sake of an internal moral principle; the latter would favor endorsing them, in the name of rationality. The conflict, then, between the two systems would result in increased response time when a participant faced a footbridge-like dilemma and made a rational judgment.

Crucially, almost all experimental studies based on Thomson's (1985) paradigm have tended to use abstract moral dilemmas, framed in such way that limits the accessibility (Kahneman, 2003) of moral utilitarian actions and consequences, asking respondents to apparently put themselves into those cognitively challenging situations.

For example:

"...The only way to save the lives of the five workmen is to push this stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die if you do this, but the five workmen will be saved."

Is it appropriate for you to push the stranger on to the tracks in order to save the five workmen?

YES/NO"

There are two striking issues in these commonly used descriptions of abstract moral dilemmas. First, although there is an explicit contextual account about the moral action and utilitarian consequences of saving the five workmen at the expense of the stranger, there is no corresponding account of saving the life of the stranger at the expense of the workmen. Hence, only 50% of the moral scenario is contextually available – a framing effect (Kahneman, 2003; Tversky & Kahneman, 1981), where different representations of outcomes make some features of the situation more accessible and others less accessible, leading to systematically different decisions. Second, the appropriateness question itself further adds to this framing effect by requiring an assessment of appropriateness on only one of the two possible moral actions ("Is it appropriate for you to push the stranger on to the tracks in order to save the five workmen"). Given the well-established role of contextual framing effects in decision-making (FeldmanHall, Mobbs, Evans, Hiscox, Navrady, & Dalgleish, 2012; Tversky & Kahneman, 1981), findings and interpretation of utilitarian moral decision-making based on these commonly used scenarios are to be treated with caution.

For the current study, in an attempt to increase the accessibility of moral utilitarian actions and consequences – utilitarian accessibility – we have developed and debiased abstract moral scenarios and questions used by researchers in psychology, experimental philosophy and neuroscience.

For example:

"....The only way to save the lives of the five workmen is to push the lone workman off the bridge and onto the track below where his large body will stop the trolley. The lone workman will die if you do this, but the five workmen

will be saved. The only way to save the life of the lone workman is not to push this workman off the bridge. The five workmen will die if you do this, but the lone workman will be saved.

Choose the option which is more appropriate for you:

Sacrifice one workman in order to save five workmen

or

Sacrifice five workmen in order to save one workman"

First, we offer a new experimental approach to study moral dilemmas by eliminating confounding variables (see, e.g., McGuire et al., 2009), allowing the footbridge dilemma to be impersonal (switching mechanism) and for the trolley dilemma to be personal (to push the worker on the track). Second, to account for utilitarian accessibility we offer presentations of moral dilemmas by using both partial textual descriptions (commonly employed in utilitarian moral research) and novel full textual descriptions of moral actions and their consequences. Third, we further reduce differences in utilitarian accessibility by offering a choice question of appropriateness, which accounts for both utilitarian alternatives (and their consequences) in moral actions (rational and irrational choice).

The results of the current study were expected to echo Green et al.'s (2001), with longer response time to personal dilemmas compared to impersonal ones, *only* when partial descriptions and a single appropriateness question were presented. In contrast, no differences in response time were expected between the dilemma types when full descriptions and a choice question were provided.

Experiment

Method

Participants.

According to power analysis with a significance level = .05, desired power = .80 and medium effect size (f^2 = .25), a total sample size of 136 was required. Participants were recruited through a recruitment service of online survey panels. A window of seven days was set for data collection; after a week had passed, 299 people (170 females, 129 males) had taken part, meeting the required sample size. Mean age was 49 years (SD = 14.07). They took part individually and received a payment of £1. All participants were treated in accordance with the ethical standards of the British Psychological Society.

Materials and Design.

Each participant was given one of eight vignettes to read, involving a moral-dilemma scenario where the type of dilemma, action involvement, task instructions and questions were manipulated. The experiment accounted for utilitarian accessibility by presenting descriptive information about the moral dilemmas: (1) by partial text description and question only (Greene et al., 2001) or (2) by full textual description and question, revealing all of the possible behavioral actions and consequences of the actions (see the supplemental materials).

An independent measures 2×2×2 design was employed, with independent variables *type of dilemma* (trolley dilemma or footbridge dilemma), *action involvement* (moral personal or moral impersonal), and *utilitarian accessibility* (partial text description and question or full text description and question [displayed information about the initial state, action, and consequences of the action]). The

dependent variables were the choice of appropriateness of action (making a rational or irrational choice), study time (reading the scenarios), and response time. Based on the consequentialist theory of moral utilitarian judgment, in this experiment we defined a rational choice as one that saves the lives of five workmen rather than of another single workman, thereby maximizing the utility of the moral action that is taken and minimizing the disutility. The order of the response options (rational and irrational) was counterbalanced across participants.

Procedure.

Instructions, scenario and question were presented in an online computer-based experiment. Participants were required to read the instructions and one moral-dilemma scenario. Then (after clicking the 'next' button) they had to choose the more appropriate option for them (binary choice between actions with rational or irrational consequences).

Results.

The effect of the independent variables on *choice*¹ was analyzed. Rational choices (choosing the option resulting in one death rather than five) were more commonly made when full information was presented and when an impersonal dilemma presented (Table 1): a logistic-regression model comprising all the main effects and interaction effects explained 38% of variance, $Rcs^2 = .38$. The main effects of accessibility (partial information vs full information), OR = 31.67, Cl.95 = [3.95; 254.08], and involvement (impersonal vs personal), OR = 0.09, Cl.95 = [0.03; 0.31], were significant, but neither the main effect of dilemma type nor any of the interaction effects were. The odds of a rational choice were 31.67 times larger when a dilemma was presented with full information than when it was presented with

reduced information. Furthermore, the odds of a rational choice were 0.09 times smaller when a dilemma involved a choice of a personal act (pushing the person) than when it involved an impersonal act (operating a switch without direct contact with the person).

Study time for a dilemma with full information was longer than when partial information was displayed; furthermore, when involvement was impersonal, time was longer than when it was personal (Table 2). A $2\times2\times2$ analysis of variance (ANOVA) showed that the main effects of accessibility (partial vs full information), F(1, 291) = 13.31, p < .001, $\varepsilon^2 = .04$, and involvement (impersonal vs personal), F(1, 291) = 5.33, p < .05, $\varepsilon^2 = .01$, were significant.

In contrast, *response time* for a dilemma with full information was shorter than when partial information was displayed (Table 3). A $2\times2\times2$ ANOVA showed that the main effect of accessibility (partial vs full information), F(1, 291) = 30.89, p < .001, $\epsilon^2 = .09$, was significant. Further analysis examined Greene and colleagues' (2001) claim that 'emotional interference' produces longer response time for emotionally incongruent responses. Specifically, the dual-process theory of moral behavior (Greene et al., 2001) predicts longer response time for a rational choice in response to a moral dilemma under the condition of personal involvement than for a rational choice under the condition of impersonal involvement. However, descriptives indicated that response time was longer for emotionally incongruent response only under the conditions of partial information, employed by Greene and colleagues (Figure 1). In support, we conducted $2\times2\times2\times2$ ANOVA, with choice rationality (response to the task) as an additional independent variable. The results show that the main effect of accessibility, F(1, 283) = 8.59, p < .01, $\epsilon^2 = .02$, and the interaction

effects of involvement by accessibility, F(1, 283) = 5.48, p < .05, $\epsilon^2 = .01$, involvement by choice rationality, F(1, 283) = 14.43, p < .001, $\epsilon^2 = .04$, and accessibility by choice rationality (rational vs irrational choice), F(1, 283) = 6.72, p < .05, $\epsilon^2 = .02$, were significant.

Follow-up simple-effect tests showed that for moral dilemmas with partial information, the interaction between involvement and choice rationality was significant, F(1, 159) = 15.60, p < .001, $\varepsilon^2 = .09$. Unsurprisingly, further simple effects within partial information revealed that the effect of choice rationality was significant, F(1, 82) = 8.69, p < .01, $\varepsilon^2 = .09$, when involvement was personal, with rational choices taking more time to make ($M_{Ln} = 2.81$; $SD_{Ln} = .38$) than irrational ($M_{Ln} = 2.16$; $SD_{Ln} = .61$); however, when involvement was impersonal, the effect was significant, F(1, 76) = 8.56, p < .01, $\varepsilon^2 = .09$, with rational choices taking less time ($M_{Ln} = 2.03$; $SD_{Ln} = .52$) than irrational ($M_{Ln} = 2.51$; $SD_{Ln} = .84$).

Yet, simple effects showed that for moral dilemmas with full information only the effect of choice rationality was significant, F(1, 138) = 10.69, p < .01, $\varepsilon^2 = .06$, with rational choices taking less time ($M_{Ln} = 1.79$; $SD_{Ln} = .49$) than irrational ($M_{Ln} = 2.19$; $SD_{Ln} = .46$). These findings suggest that any emotional interference, with rational choices taking more time to make, appears as an artifact of presenting partial information and disappears when full information is presented, with rational choices taking less time.

Discussion

Our results reveal that variation in utilitarian accessibility produces variation in moral choices. In particular, displaying full information regarding moral actions and

consequences resulted in an increase of rational choices. Moreover, the effect of utilitarian accessibility was general in that it occurred across types of involvement (both personal and impersonal) and types of dilemma (both trolley and footbridge). Previous research (e.g., Greene et al., 2001) found that judges took more time to judge an action as rational when a moral dilemma was personal. However, type of dilemma and involvement were confounded (McGuire et al., 2009), and utilitarian accessibility was not manipulated. Accordingly, our results show that response time for a dilemma with full information was shorter than when partial information was displayed. However, neither the main effects of the manipulated variables involvement and dilemma, nor the interaction effects influenced response time.

We further examined Greene and colleagues' (2001) claim that 'emotional interference' produces longer response time for emotionally incongruent responses. This prediction was only confirmed when participants make a rational choice in response to a moral dilemma under the condition of personal involvement (e.g., judging it appropriate to push the man off the footbridge in the footbridge dilemma) under conditions of partial information. In contrast, with full information presented, rational choices were made faster. Therefore, our results suggest that any emotional interference, with rational choices taking more time to make, is an artifact of presenting partial information and does not happen when full information is presented, with rational choices taking less time. Given our results, a more plausible interpretation of increased response time with rational answers under conditions of partial information is reduced utilitarian accessibility rather than 'emotional interference'. When decision-makers are presented with full contextual information about a particular moral action and its consequences, the framing effect will be eliminated and mental simulation will not entertain other possible outcomes of the

scenario (e.g., FeldmanHall et al., 2012). Therefore, decision-makers are more vividly confronted with the effect of the action (whether personal or impersonal). It is plausible that limited utilitarian accessibility of moral actions and consequences results in a psychological uncertainty and corresponding mental simulations (compensating for reduced accessibility of moral actions and consequences). In contrast, comprehensive information about moral actions and consequences may eliminate uncertainty, and boost utility maximization in moral choices, with rational choices taking less time. Such an interpretation might be accommodated by 'situation models' (e.g., Glenberg, Meyer, & Lindem, 1987), in which linguistic descriptions are understood by simulating perceptual and motor aspects of those descriptions. There is empirical support for this simulation view that also demonstrates that switching between simulations takes time both when switching perceptual modality (Pecher, Zeelenberg, & Barsalou, 2003) and semantic interpretation (Ferguson, Tresh, & Leblond, 2013).

Our main finding is the effect of utilitarian accessibility on judgment of appropriateness and response time. Therefore, we agree with McGuire et al.'s (2009) recommendation that "More research needs to be done at a behavioral level in order to fine-tune the questions being asked before work identifying the neural correlates of moral decision-making can be useful" (p. 580).

References

- Bentham, J. (1789/1970). *An Introduction to the Principles of Morals and Legislation.*New York: Haffner Press.
- FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: the relationship between real and hypothetical moral choices. *Cognition*, *123*(3), 434-441.
- Forbes, C. E., & Grafman, J. (2010). The Role of the Human Prefrontal Cortex in Social Cognition and Moral Judgment. *Annual Review of Neuroscience*, 33, 299-324.
- 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.
- Ferguson, H. J., Tresh, M., & Leblond, J. (2013). Examining mental simulations of uncertain events. *Psychonomic Bulletin & Review, 20 (2),* 391-399
- Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, 26, 69–83.
- 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, 2105-2108.
- Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, *6*(12), 517-523.
- 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.
- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814-834.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58* (9), 697–720.
- Kant, I. (1959). *Foundation of the metaphysics of morals*. (Lewis White Beck, Trans.) Indianapolis: Bobbs-Merrill. (Original work published 1785).
- 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*, *45*, 577–580.
- Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future.

 Trends in Cognitive Sciences, 11(4), 143-152.
- Mikhail, J. (2009). Moral grammar and intuitive jurisprudence: a formal model of unconscious moral and legal knowledge. *Psychology of Learning and Motivation*, *50*, 27-100.
- Moll, J., de Oliveira-Souza, R., & Eslinger, P. J. (2003). Morals and the human brain: a working model. *Neuroreport*, *14*(3), 299-305.
- 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.
- Nakamura, K. (2013). A closer look at moral dilemmas: latent dimensions of morality and the difference between the trolley and footbridge dilemmas. *Thinking and Reasoning*, 2, 178-204.

- Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychological Science*, *14*, 119–124.
- Simon, H. A. (1968). The future of information processing technology. *Management Science*, *14*, 619-624.
- Thomson, J. J. (1985). The trolley problem. In J. M. Fischer & M. Ravizza (Eds.), *Ethics: Problems and Principles*. Fort Worth, TX: Harcourt Brace Jovanovich.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*(6), 476–477.

Acknowledgments

Petko Kusev and Paul van Schaik are supported by Economic and Social Research Council Grant RES-000-22-1768. We also thank the Nuffield Foundation (SGS36177) and The British Academy (SG47881/SG091144) for supporting Petko Kusev in his research.

Footnote

¹ Irrational choice was the reference category and rational choice as the response category.

Tables

Table 1
Choice as a function of involvement, accessibility and dilemma type

		Trolley		Footbridge	
Involvement	Accessibility (information)	Irrational	Rational	Irrational	Rational
Impersonal	Partial	6%	7%	8%	5%
		(19)	(21)	(23)	(14)
	Full	0%	12%	2%	9%
		(1)	(35)	(6)	(28)
Personal	Partial	13%	1%	12%	1%
		(39)	(4)	(36)	(4)
	Full	2%	10%	4%	8%
		(5)	(29)	(11)	(24)

Note. Figures are percentages with frequencies in brackets.

Table 2

Descriptives for study time by involvement, accessibility and dilemma type

Involvement	Accessibility (information)	Tr	Μ	SD	Μ	SD
Impersonal	Partial	Т	3.23	0.50	3.29	0.60
		U	28.43	14.27	31.38	17.28
	Full	Т	3.40	0.53	3.43	0.54
		U	34.55	20.65	36.46	27.16
Personal	Partial	Т	3.12	0.44	3.15	0.46
		U	21.63	8.74	25.56	9.85
	Full	Т	3.36	0.47	3.35	0.50
		U	30.01	12.49	32.10	16.86

Note. The frequency distribution of study time was positively skewed and this was considerably improved by logarithmic transformation.

Tr: transformation. T: logarithmically transformed. U: untransformed (original).

Table 3

Descriptives for response time by involvement, accessibility and dilemma type

			Trolley		Footbridge	
Involvement	Accessibility (information)	Tr	М	SD	М	SD
Impersonal	Partial	Т	2.28	0.74	2.30	0.76
-		U	13.43	13.88	13.15	9.57
	Full	Τ	1.85	0.52	1.89	0.54
		U	7.25	3.92	7.62	4.14
Personal	Partial	Τ	2.16	0.60	2.29	0.63
		U	10.51	7.25	12.19	8.76
	Full	Τ	1.85	0.50	1.86	0.50
		U	7.15	3.53	7.25	3.48

Note. The frequency distribution of study time was positively skewed and this was considerably improved by logarithmic transformation.

Tr: transformation. T: logarithmically transformed. U: untransformed (original).

Figure Captions

Figure 1. Mean response time as a function of accessibility, involvement, and choice rationality (time in seconds).

Figure 1.



