Synthese
DOI 10.1007/s11229-012-0121-z

Use of a Rasch model to predict response times to utilitarian moral dilemmas

Jonathan Baron · Burcu Gürçay · Adam B. Moore · Katrin Starcke

Received: 1 December 2011 / Accepted: 30 April 2012

Abstract A two-systems model of moral judgment proposed by Joshua Greene holds that deontological moral judgments (those based on simple rules concerning action) are often primary and intuitive, and these intuitive judgments must be overridden by reflection in order to yield utilitarian (consequence-based) responses. For example, one dilemma asks whether it is right to push a man onto a track in order to stop a trolley that is heading for five others. Those who favor pushing, the utilitarian response, usually take longer to respond than those who oppose pushing. Greene's model assumes an asymmetry between the processes leading to different responses. We consider an alternative model based on the assumption of symmetric conflict between two response tendencies. By this model, moral dilemmas differ in the "difficulty" of giving a utilitarian response and subjects differ in the "ability" (tendency) to give such responses. (We could just as easily define ability in terms of deontological responses, as the model treats the responses symmetrically.) We thus make an analogy between moral dilemmas and tests of cognitive ability, and we apply the Rasch model, developed for the latter, to estimate the ability-difficulty difference for each dilemma for each subject. We apply this approach to five data sets collected for other purposes by three of the co-authors. Response time (RT), including yes and no responses, is

J. Baron (✉)
Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104-6241, USA
e-mail: baron@psych.upenn.edu

B. Gürçay
Department of Psychology, University of Pennsylvania, Pennsylvania, PA 19104-6241, USA

A. B. Moore
Georgia Institute of Technology, Atlanta, GA, USA

K. Starcke
Universität Duisburg-Essen, Duisburg, Germany

Published online: 13 June 2012

longest when difficulty and ability match, because the subject is indifferent between the two responses, which also have the same RT at this point. When we consider yes/no responses, RT is longest when the model predicts that the response is improbable. Subjects with low ability take longer on the "easier" dilemmas, and vice versa.

Keywords Moral judgment · Rasch model · Dual system

1 Introduction

A type of research on moral judgment is concerned with responses to brief moral dilemmas that pit utilitarian consequences against moral rules (Spranca et al. 1991, review early studies). For example, should you push a man off of a footbridge, to his death, if he would land in the path of a runaway trolley and thus prevent it from killing five other people working on the tracks? The consequences are one death versus five, but people recoil at the idea of pushing a man to his death and typically say that they should not do this. They often say that they are following a rule against direct killing of innocent people.

It has become apparent that some dilemmas are "easier" than others, in the sense that they elicit many more utilitarian responses and seem to be less emotional and conflicting. For example, if the choice is to switch the trolley to another track heading for two other people, rather than push the man, most people say they should do it, even though they would directly cause more deaths than by pushing the man.

Greene et al. (2001) argued that moral dilemmas varied in terms of how emotionally engaging they were, and that both reason and emotions affected moral judgments. Consistent with a great deal of theorizing elsewhere (e.g., Epstein 1994; Finucane et al. 2000; Hammond 1996; Kahneman and Frederick 2002; Sloman 1996; Stanovich and West 2002), Greene and his colleagues have proposed an account in terms of two systems. The first is automatic, intuitive, sensitive to emotion, and fast. It yields an immediate reaction to the emotional aspects of the dilemma and leads to judgments favoring simple moral rules. The second is reflective, slower, and more sensitive to the consequences. In the footbridge problem, the initial intuitive response says not to push, but some subjects think again and change their minds, after thinking about the consequences. By contrast, the initial intuitive response to the switching problem is either absent or favors switching the trolley. But the model generally treats the two systems as asymmetrical with respect to response times (RTs). The two-systems account in general is supported by functional magnetic resonance studies showing that different brain regions are activated when people are presented with more emotionally engaging moral dilemmas. Other support comes from studies of the effects of time pressure and individual differences (Greene et al. 2008). We do not deal with all of these issues here, only the evidence from RTs.

Greene and his colleagues suggested that dilemmas that evoke the system-1 emotional response more strongly are more "personal", which they have defined as "(i) likely to cause serious bodily harm, (ii) to a particular person, (iii) in such a way that the harm does not result from the deflection of an existing threat onto a different party…A moral violation is impersonal if it fails to meet these criteria" (Greene 2005;

Greene and Haidt 2002, p. 519). Greene et al. (2001) also argued that an important feature of footbridge-like (personal) dilemmas was that they elicit strong emotional reactions, and his hypothesis was supported by his findings showing more activation in emotion related brain areas (Greene et al. 2004). Greene (2009) stated that the personal/impersonal distinction was just a preliminary theory, and thus a stand-in, until a better explanation can be put forward; and later on Greene et al. (2009) argued that what made people judge it morally acceptable to sacrifice one person's life in order to save others' in some cases and not in others was an interaction between personal force applied on the sacrifice and intention. In this paper we use the old terminology, in line with other studies that continue to make this distinction (e.g., Moore et al. 2011).

Gürçay and Baron (in preparation, henceforth GB2011) tested Greene's two-systems model account of moral judgment in a different way. In each of two studies, GB2 and GB3, In each of the two studies, GB2 and GB3, subjects saw moral dilemmas very similar to those used by Moore et al. (2008), which were based on those used in Greene et al. (2001, 2004) but modified to equate reading times and eliminating some dilemmas. GB2 and GB3 examined the personal/impersonal distinction, and the effects of asking subjects to respond slowly and reflectively versus fast and intuitively. The speed manipulation was designed to test the hypothesis that the slow-reflective condition would yield more utilitarian responses, especially for the personal dilemmas, where the intuitive response was less likely to be utilitarian. GB2 and GB3 found no effect of the manipulation on the number of utilitarian responses. The manipulation affected RT, as it should have, but not differentially for yes and no responses.

Differences in utilitarian responding are determined not only by the dilemmas used but also by the individual. In the typical moral judgment tasks, individual differences are typically large, easily measured, and highly reliable even with just a few items (e.g., Baron and Ritov 2009; Greene et al. 2008). This result suggested that it would be worthwhile to analyze choices and times in terms of subject differences as well as dilemma differences.

If we put this fact together with the differences among dilemmas, we are led to think of a moral-judgment task as analogous to an ability test such as a test of vocabulary. Correctly defining a word in a vocabulary test is analogous to giving the utilitarian response to a dilemma. People differ in the size of their vocabulary, and words differ in their difficulty, which is roughly the proportion of people who know them.

A simple mathematical model of this situation was proposed by Rasch (1961) and has been widely used with various modifications (Lord and Novick 1968). Here we use the simple form because of its advantage as a summary of data for our purposes (hence not because it fits the data better than other forms). The model assigns to each dilemma a number representing its "difficulty," and to each subject a number representing her "ability" (to produce utilitarian responses). The model then states that the probability of a utilitarian answer is a logistic function of the *difference* between ability and difficulty. Specifically, $p = \frac{e^{\beta_n - \delta_i}}{1+e^{\beta_n - \delta_i}}$, where p is the probability of a "correct" (utilitarian) answer, β_n is the ability of subject n and δ_i is the difficulty of item i. When ability and difficulty are equal, so that the difference $\beta_n - \delta_i$ is 0, the probability is 0.50. As the difference increases so that ability is higher, the probability increases, with the

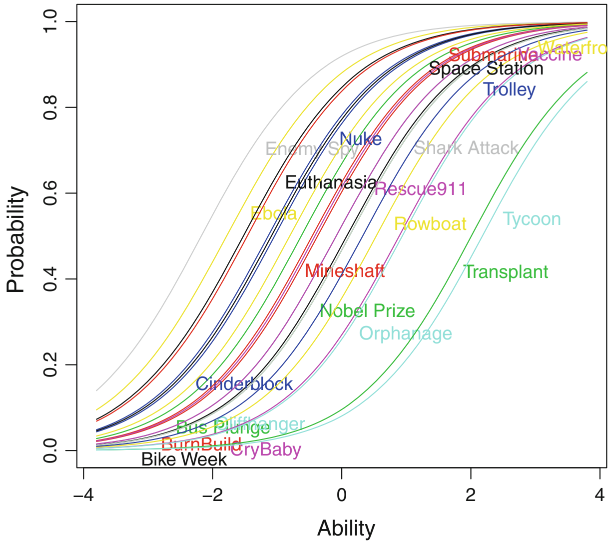

Fig. 1 Best fitting Rasch *curves* for data from Moore et al. (2008, abbreviated as M2008), showing the probability of a utilitarian answer as a function of the subject's ability. Each *curve* represents a dilemma. (*Colors* are arbitrary.) Some dilemmas are *labeled*; the position of the labels is not meaningful. (Color figure online)

increase slowing as the probability approaches 1.00; and conversely for decreasing difference.

We fit the model using the ltm package in R (Rizopoulos 2006). Figure 1 illustrates best fitting curves to the data from Moore et al. (2008, henceforth M2008). Each curve represents a dilemma, showing the probability of a yes response as a function of the subject's ability. The dilemmas differ in difficulty, so that both the dilemma and the subject's ability determine the probability of a yes answer.

The model assumes, incorrectly (as we discuss below), that all items are equal in their power to discriminate subjects on the basis of ability, that subjects do not say yes to more difficult dilemmas when they have said no to easier ones, and so on. But our primary interest in the model is to estimate, for each subject, the point where difficulty and ability are equal, and the model fit well enough for this purpose. The items near this point in difficulty should have a response probability of about 0.50 and should thus be the most difficult for a given subject, yielding the slowest RTs.

In Greene's studies, dilemmas differed in the "difficulty" for giving utilitarian judgments; some were quite easy and some were difficult, perhaps because they were "personal" but perhaps also for other reasons. Each subject would therefore find some dilemmas more conflicting than others, but the particular dilemmas would differ depending upon the subject's "ability" (tendency) to make utilitarian judgments. Thus, the fact that specific dilemmas are more conflicting and might cause subjects to take longer in making a judgment can be a consequence of the matching of subjects and dilemmas. Highly utilitarian (high-ability) subjects would be conflicted only by the most difficult items. For subjects who are much more deontological, with lower ability (to make utilitarian judgments), the situation might reverse, making the easy cases, like the simple trolley problem with switching, more difficult for them. Given the usual sample of dilemmas in previous studies, most subjects seem to be on the relatively high-ability side, leading to slower responses only for the most difficult dilemmas, which are mostly the "personal" ones.

In this article, we examine RT as a function of ability and difficulty, as determined from fitting the Rasch model to each study, to five studies, two from GB2011, one from M2008, one from Moore et al. (2011, henceforth M2011), and one from Starcke et al. (2012, henceforth S2012). Our main hypothesis is that RTs in general are longest when the dilemma is difficult for the subject, given the subjects inclination to respond in terms of consequences (i.e., ability). By difficult, we mean that ability and difficulty are equal, so the prediction is that the subject has a 50 % chance of responding "yes" (in favor of the utilitarian response).

We also show RTs for "yes" and "no" responses separately. Greene's model might predict that, at the point of equality, RTs for "yes" responses would take longer, because the "no" response is typically generated by system-1. And Greene's model predicts that utilitarian responses to difficult dilemmas have high RT because of the need to invoke system-2 to over-ride the initial response. We propose, alternatively, that these high RT responses are consistent with the general finding that low-probability choices have high RT (e.g., Petrusic and Jamieson 1978), a phenomenon that is consistent with several models of choice RT based on the idea of gradual accumulation of statistically variable information supporting one response or the other over time (e.g., Busemeyer and Townsend 1993; Link 1992; Ratcliff and McKoon 2008): when information supporting one option is weak, the option is unlikely to be chosen and, when it is chosen, more time is required to accumulate sufficient information.

The main point is that we can view RTs to moral dilemmas as typical of responses under conflict between two options, with some considerations pulling one way and some the other way. When these considerations are balanced, the decision will be difficult and take longer. When they are unbalanced, one response will be more probable and faster. And this prediction is symmetric with respect to the two responses.

2 Method

For all the data, we reduced the data sets to the following variables: subjects, dilemmas, response (1 = utilitarian, 0 = non-utilitarian), order (the order in which the dilemma was presented),[1] and normalized log RT. That is, RT represented the log of the RT, minus each subject's mean log RT. Using the Rasch function in ltm, we fit the Rasch model and obtained a difficulty measure for each dilemma and an ability measure for each subject. The Rasch modeling used only response (1 or 0), subject, and dilemma.[2] In all studies, we omitted subjects who made the same response to every dilemma, and (in S2012) the single dilemma that yielded no positive (yes) responses.

[1] In all studies, RT declined with order, and in three of the five studies, utilitarian responding increased significantly with order.

[2] For GB2 and GB3, subjects responded five times in immediate succession to each dilemma, with varying numbers of harms. Most responses to each dilemma were the same; the number of harms had little effect. We thus took the majority of responses to indicate yes (utilitarian) or no.

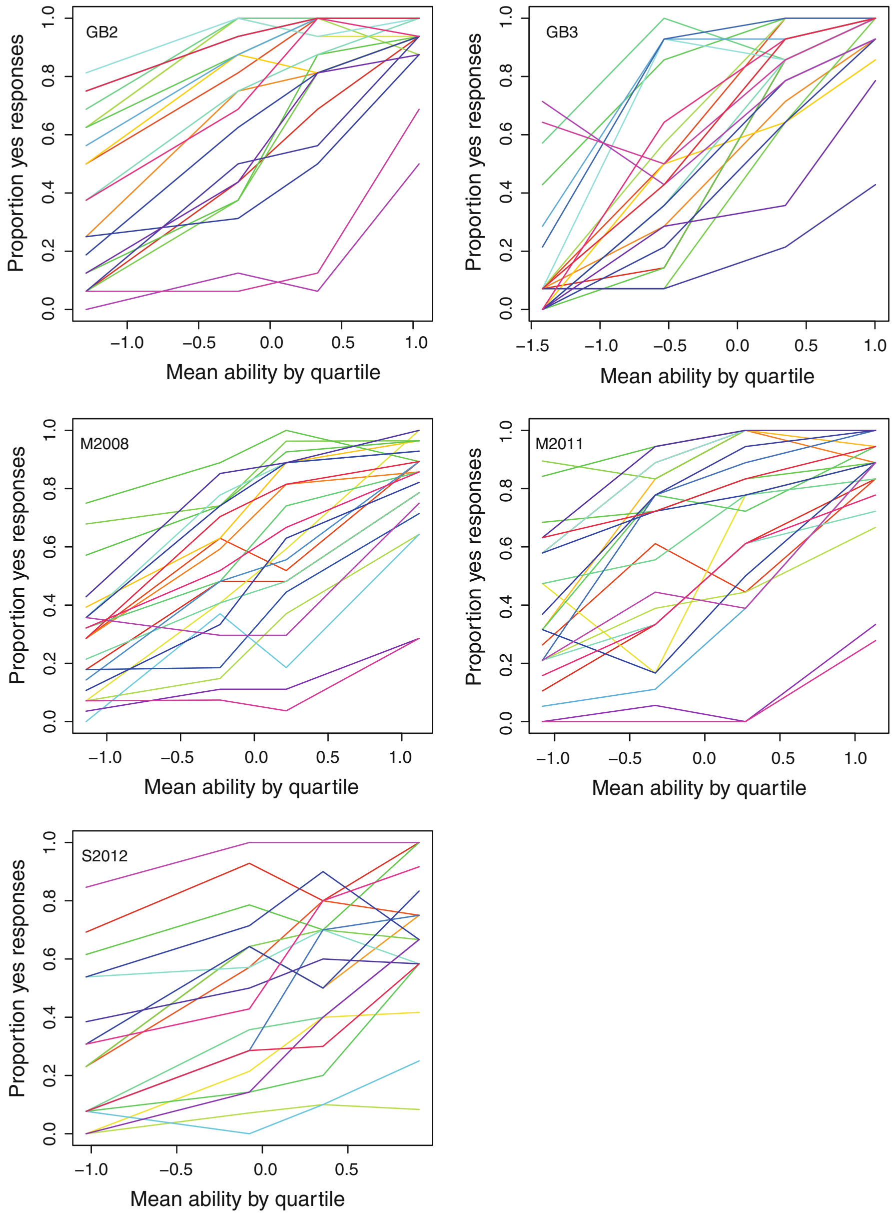

Fig. 2 Proportion of yes responses as a function of mean ability of each quartile of ability. Each *line* represents a dilemma

Figure 2 shows approximately the data on which the Rasch models were fit—approximate because ability scores are grouped by quartile, and the plots display proportions as a function of mean ability score for each quartile. It is apparent that the form of these lines differs from the ideal Rasch model shown in Fig. 1, in several ways. Some of these can be understood as "random noise", e.g., when the functions are not monotonically increasing. Other deviations are typical of other sorts of data that have led to revisions in the Rasch model itself: dilemmas differ in their slopes even at their midpoints; they seem to approach upper asymptotes less than 1 or lower asymptotes greater than 0. We could have fit the data better with a model that allowed for such deviations, but these more complex models do not allow a clear definition of ability and difficulty, and our purpose here was to derive these measures so that we could examine their effects on RT, especially at the point where they were equal. The Rasch model is thus somewhat of a Procrustean bed, a model that is false but useful as a description of data. However, it does seem that the items in each study are measuring a single dimension of individual differences fairly well. Coefficient α's for the studies were, respectively: GB2, 0.88; GB3, 0.95; S2012, 0.60; M2008, 0.82; M2008, 0.82.

We tested three hypotheses statistically. The first hypothesis was that RT (logged and normalized by subject) should be highest when ability is equal to difficulty. We tested this with a mixed model in which RT was predicted from the square of the difference (ability minus difficulty, squared), ability, difficulty, and order, with crossed random effects for subjects and dilemmas. We used the lmer function in the lme4 package of R (Baayen et al. 2008; Bates 2005; Bates et al. 2011; R Development Core Team 2011), along with the languageR package (Baayen 2011), for assessing "*p* values" with Markov Chain Monte-Carlo sampling.[3] The hypothesis predicts a negative coefficient for $(\textit{ability-difficulty})^2$; the peak RT should occur when the difference is 0. This hypothesis is predicted by our model even if individuals do not differ much.

The second hypothesis is the more interesting prediction that some people will take longer on difficult dilemmas and others will take longer on easy dilemmas. To test this hypothesis, we computed the correlation across subjects between two measures for each subject. One was the subject's ability. The other was the within-subject slope of (logged) RT regressed on difficulty. Subjects with high ability should have longer RTs on the more difficult items, so the within-subject correlation should be more positive for them. And subjects with low ability should have longer RTs for the easy dilemmas, so the within-subject correlations should be more negative for them. Thus, we hypothesize a positive correlation between subject ability and the within-subject correlation between difficulty and RT.

The third hypothesis is that this positive correlation leads to opposite results for subjects with high and low ability. We define high-ability subjects as those with Rasch ability scores above 0.5, and low ability with scores below -0.5. These subjects were approximately the top and bottom quartile. The high-ability subjects should be slower on the difficult items, and the low-ability subjects should be faster on the easier items. We test this with t-tests on the mean slopes (of RT on difficulty) for the two groups of subjects.

We also examine, graphically (in Fig. 3), how RT depends on yes and no responses. Of particular interest is whether RT for yes and no is the same when ability and difficulty are equal, the point at which the two responses should be equally likely. We also illustrate graphically the increased RT for low-probability responses of either type.

[3] For example: $lmer(RT\ I((\textit{ability-difficulty})^2) + \ ability + difficulty + item + (1|dilemma) + (1|subjects))$. Item is the order in which the dilemma was presented.

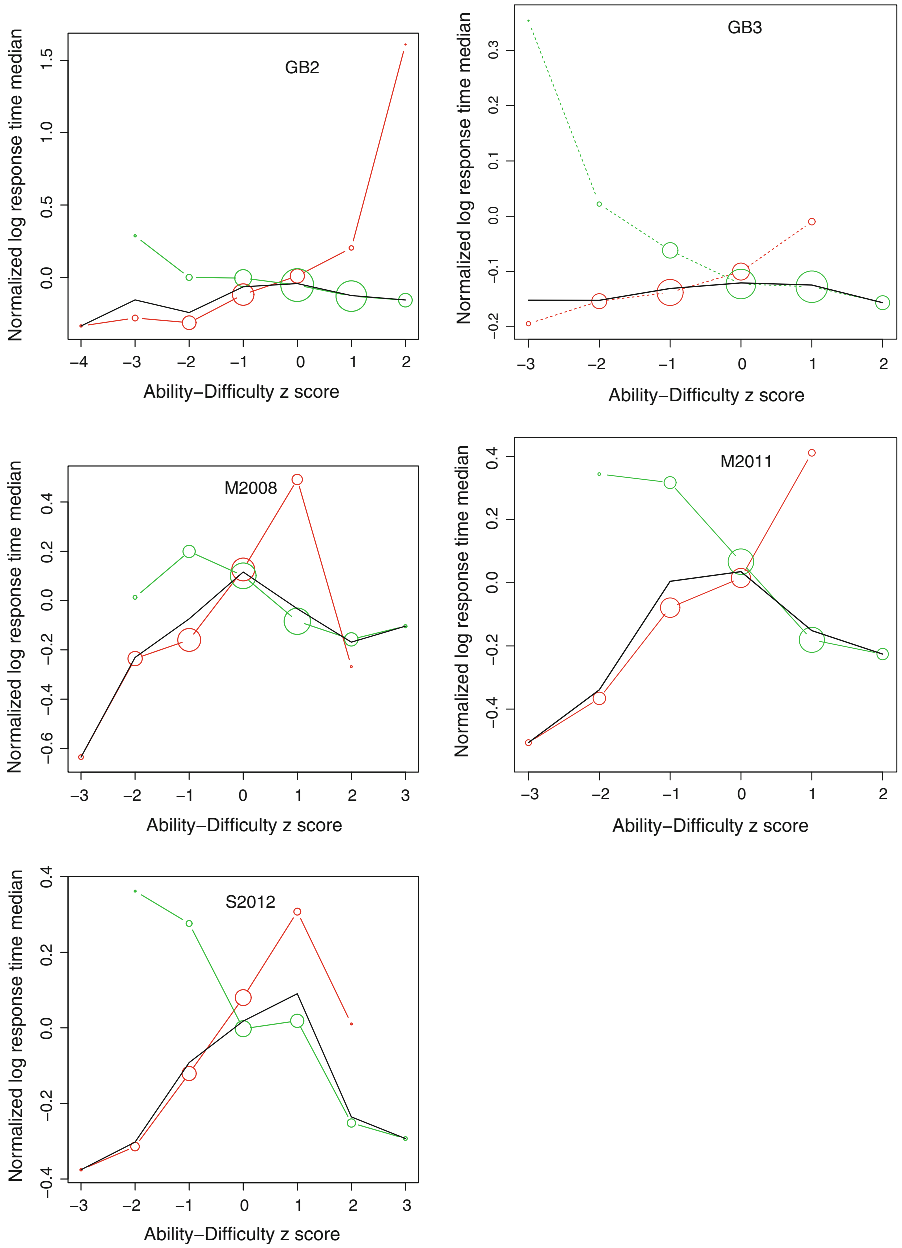

Fig. 3 Mean log normalized RT as a function of ability–difficulty (*scaled* and *rounded*) for the five studies. *Circle areas* are proportional to the number of observations in each point; *red* indicates "no", *green*, "yes", *black* is both combined. (Color figure online)

Table 1 Main results for the five studies

	BG2	BG3	s2012	M2008	M2011
Coefficient of (ability − difficulty)^2	−0.018	−0.025	−0.019	−0.031	−0.040
MCMC p value	0.0001	0.0036	0.0001	0.0001	0.0001
r(ability, RT coefficient on difficulty)	0.45	0.43	0.59	0.44	0.62
p	0.0002	0.0009	0.0000	0.0000	0.0000
Coefficient for ability >0.5	0.04	0.13	0.03	0.07	0.10
p	0.0388	0.0014	0.0577	0.0283	0.0010
Coefficient for ability <-0.5	−0.05	−0.10	−0.05	−0.09	−0.11
p	0.0116	0.0352	0.0108	0.0002	0.0004

3 Results

Table 1 shows the main results, which are consistent across the five studies. In all studies, the quadratic term for predicting RT from ability minus difficulty was negative, as hypothesized, and significant (first row of Table 1). Figure 3 shows plots of the relevant RT data, as medians of z-scores of log RT. The heavy solid line shows the main test of the first hypothesis, the peak when ability and difficulty are equal, including both yes and no responses. In general, overall RT is greatest when ability and difficulty are equal.

The dashed lines show the RTs for yes and no responses, with the areas of the circles indicating the number of responses at each point, hence the probability. Low probability responses "against type"—that is, negative responses when ability > difficulty positive when ability < difficulty—take longer. Note that, in GB2 and GB3, the increase in RT for these rare responses is so great that the scale of the plot is increased and the curvature of the solid line is barely visible, despite being highly significant. Importantly, yes and no RTs are about equal when ability = difficulty, contrary to the hypothesis that systems-1 and -2 are asymmetric with respect to RT.

The other results in Table 1 show the effect of individual differences in ability. In all studies ability correlates with the relation between RT and difficulty (second row). And, in all studies, consistent with our third hypothesis, high-ability subjects take longer on the difficult dilemmas (third row) and low-ability subjects take longer on the easy dilemmas (fourth row).

4 Discussion

We found that people differ considerably, so that some people have difficulty deciding and thus take longer on "easy" dilemmas where most people choose the utilitarian response and other people have difficulty and take longer on "hard" dilemmas. In general, we find no asymmetry of the sort that might be predicted by the two-system model.

These results imply that previous findings concerning individual differences and group differences in responses and RT may depend on the selection of dilemmas and subjects. In particular, the evidence to date supports the view that responses to personal dilemmas involve system-2 reflection after an initial intuitive non-utilitarian response (whether or not this reflection actually overrides the initial response). But personal dilemmas are more difficult, in our sense. For subjects with "low ability", these dilemmas might be "easier" and have faster RTs. For these subjects, slower times occur on the easier dilemmas that evoke utilitarian responses in most people.

We also note that the dilemmas we used, all derived from those used originally by Greene and his colleagues, were selected in part because, for many of them, the utilitarian response was put in conflict with an intuitive moral rule, such as a rule against causing direct harm. It might be possible to find cases in which the moral rule is itself non-intuitive and the source of conflict (Baron 2011), as in cases in which a rule dictates excessively harsh punishment (such as excommunication for allowing an abortion to save the life of a pregnant woman).

Our results do not dispute the two-systems theory in general. Long RTs may involve use of the reflective system. Reflection itself may be the result of conflict, of a failure to decide quickly between the two response options, rather than the result of a particular direction of the intuitive response. It may, however, be difficult to distinguish reflection from what ordinarily happens when the two response options are equally attractive (or equally unattractive), as in psychophysical tasks when comparative judgments involve stimuli that are very similar.

References

Baayen, R. H. (2011). *languageR*: *Data sets and functions with* "*Analyzing Linguistic Data*: *A practical introduction to statistics*". R package version 1.4. http://CRAN.R-project.org/package=languageR. Accessed 1 Dec 2011.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Baron, J. (2011). Utilitarian emotions: Suggestions from introspection. *Emotion Review* (special issue on "Morality and emotion" edited by Joshua Greene)*, 3*, 286–287.

Baron, J., & Ritov, I. (2009). Protected values and omission bias as deontological judgments. In D. M. Bartels, C. W. Bauman, L. J. Skitka & D. L. Medin (Eds.), *Moral judgment and decision making*, Vol. 50 in B. H. Ross (series editor), *The psychology of learning and motivation* (pp. 133–167). San Diego, CA: Academic Press.

Bates, D. M. (2005). Fitting linear mixed models in R. *R News, 5*, 27–30.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4*: *Linear mixed-effects models using S4 classes*. R package version 0.999375-42. http://CRAN.R-project.org/package=lme4. Accessed 1 Dec 2011.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision Field Theory: A dynamic cognition approach to decision making. *Psychological Review, 100*, 432–459.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49*, 709–724.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making, 13*, 1–17.

Greene, J. (2005). Cognitive neuroscience and the structure of the moral mind. In S. Laurence, P. Carruthers, & S. Stich (Eds.), *The innate mind: Structure and contents* (pp. 338–352). New York: Oxford University Press.

Greene, J. D. (2009). Dual-process morality and the personal/impersonal distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology, 45*, 581–584.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences, 6*, 517–523.

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. 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*, 389–400.

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*, 1144–1154.

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*, 364–371.

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York: Oxford University Press.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Link, S. W. (1992). *The wave theory of difference and similarity*. Cambridge, MA: Psychology Press.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science, 19*, 549–557.

Moore, A. B., Lee, N. Y. L., Clark, B. A. M., & Conway, A. R. A. (2011). In defense of the personal/impersonal distinction in moral psychology research: Cross-cultural validation of the dual process model of moral judgment. *Judgment and Decision Making, 6*, 186–195.

Petrusic, W. F., & Jamieson, D. G. (1978). Relation between probability of preferential choice and time to choose changes with practice. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 471–482.

R Development Core Team. (2011). *R*: *A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org. Accessed 1 Dec 2011.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. IV, pp. 321–334). Berkeley, CA: University of California Press.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*, 873–922.

Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses. *Journal of Statistical Software, 17*, 1–25. URL http://www.jstatsoft.org/v17/i05/. Accessed 1 Dec 2011.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22.

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgmentand choice. *Journal of Experimental Social Psychology, 27*, 76–105.

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate?. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 421–440). New York: Cambridge University Press.

Starcke, K., Ludwig, A., & Brand, M. (2012). Anticipatory stress interferes with utilitarian moral judgment. *Judgment and Decision Making, 7*, 61–68.