

## **UC Merced**

### **Proceedings of the Annual Meeting of the Cognitive Science Society**

#### **Title**

Moral Reasoning as Probability Reasoning

#### **Permalink**

<https://escholarship.org/uc/item/5gn4v1fm>

#### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

#### **Authors**

Shou, Yijun

Song, Fei

#### **Publication Date**

2015

Peer reviewed

## Moral Reasoning as Probability Reasoning

Yiyun Shou (yiyun.shou@anu.edu.au)

Research School of Psychology, The Australian National University  
Canberra, ACT, Australia

Fei Song (sfei618@gmail.com)

Department of Philosophy, The University of Hong Kong, Hong Kong

### Abstract

Previous studies found that the likelihood of subjects to choose a deontological judgment (e.g., allowing harm) or a consequentialist judgment (e.g., doing harm) varied across different moral dilemmas. The present paper explored if the variation can be explained by the differentiation of the perceived outcome probabilities. We generated moral dilemmas that were similar to the classical trolley and footbridge dilemmas, and investigated the extent to which subjects were sensitive to the outcome probabilities. Results indicated that the majority of subjects, including both those who initially chose a deontological decision and those who initially chose a consequentialist decision could be sensitive to outcome probabilities. The likelihood of being sensitive to the probabilities was invariant across different dilemmas. The variation of the choice behaviors across different dilemmas might be associated with the variation of the estimated outcome probabilities.

**Keywords:** probability judgment; moral reasoning; moral dilemma

### Introduction

Moral reasoning has been under long-term intellectual scrutiny. Recent psychological investigations of moral reasoning frequently employ moral dilemmas that render conflicts between moral requirements (Crockett, 2013). Moral dilemmas commonly engender conflict between two major type of moral reasoning: deontological and consequentialist moral reasoning. A deontological moral judgment primarily concerns the actions *per se*, that is, whether it is consistent with moral principles, rules or duties. On the other hand, a consequentialist judgment primarily concerns the outcome of each possible action and aims to choose the one with the best outcome.

The trolley dilemma requires people to decide between killing an innocent individual and allowing five innocent to be killed. The former is often considered as a consequentialist decision, under which the loss is minimized. In contrast, allowing five to die is taken as deontological decision under which the action of killing is regarded as a deontological violation. Crockett (2013) associated consequentialist reasoning with a model based system, in which the reasoner starts from the current action, searches through the decision tree and evaluates the best outcome of the action. In contrast, the deontological reasoning is associated with the model-free

evaluation, where the current action and forward searching are not activated.

Recent research suggests that the types of moral reasoning may be shaped by the interaction and competition between two distinct psychological systems: an automatic emotion process and a controlled conscious reasoning process (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Paxton, Ungar, & Greene, 2012). Greene et al.(2001) argued that a deontological decision might be driven by emotional arousal, while consequentialist reasoning is the result of the controlled reasoning process.

Numerous studies have shown that the majority of people perceive the consequentialist choice as morally preferred option in the trolley dilemma (Crockett, 2013). However, controversial findings rose in different variants of moral dilemmas. For instance, the footbridge dilemma—in which one needs to decide between pushing a fat man over bridge and allowing five people to die—yielded distinctive decision patterns (Lerner, Li, Valdesolo, & Kassam, 2014; Valdesolo & DeSteno, 2006). The proportion of subjects who preferred the consequentialist choice over the deontological choice can vary case by case (Cummins & Cummins, 2012). These studies suggest that some case-relevant features might influence the decision making process.

Greene made a distinction between a personal dilemma like the trolley dilemma and an impersonal dilemma like the footbridge dilemma. The personal dilemma triggers the negative response to a harmful act which treats an agent as the only means to an end, whereas the impersonal dilemma fails to trigger negative response to a harmful act which is only a side effect (Greene, Nystrom, Engell, Darley, & Cohen, 2004). It has been systematically found that personal dilemmas commonly produce more deontological judgments among subjects, while impersonal dilemmas commonly produce consequentialist judgments (Cummins & Cummins, 2012; Moll & de Oliveira-Souza, 2007).The personal dilemmas, featuring the involvement of the physical contact, may trigger higher emotional arousal, which in turn result in the higher likelihood among subjects to choose deontological-like judgments.

Greene et al (2001) presented neuroimaging evidence and showed that the footbridge dilemma was associated with greater activity in emotion associated brain areas such as the posterior cingulate gyrus (Brodmann Area 23/31) and bilateral angular gyrus (Brodmann Area 39). Greene (2009) implied that the physical contact may induce emotional arousal, which promotes subjects to be more likely to engage

in deontological reasoning. In fact these two brain areas are not restricted to emotion relevant processes. For example, the posterior cingulate gyrus was also found to be associated with the cognitive process involved in the evaluation of the values of choices (Rushworth & Behrens, 2008), while the bilateral angular gyrus was found to be activated during decision making under uncertainty (d'Acremont, Fornari, & Bossaerts, 2013).

An alternative explanation for the higher likelihood of subjects' preferences for allowing harm in personal dilemmas like the footbridge case can be that the decision of allowing harm can be a result of either endorsing deontological reasoning or endorsing consequentialist reasoning. The consequentialist reasoning recruits the model-based evaluation while the deontological reasoning recruits the model-free evaluation. To avoid confusion, we distinguish the two reasoning types from the two choices observed in a moral dilemma. We name the choice of "doing harm" as consequentialism-like choice (CLC), and the choice of "allowing harm" as the deontology-like choice (DLC). The DLC that allows five people to die can be perceived as justifiable via either types of reasoning. A reasoner who adopts consequentialist reasoning can make a DLC when the perceived utility of doing harm is lower than the utility of allowing harm. The perceived utility can be altered by the reasoner's probability estimates of the outcomes given the two choices.

Most previous studies did not explicitly indicate how likely the outcomes would occur given that each action had been taken when presenting the moral dilemmas. Subjects may estimate the outcome probabilities based on their prior knowledge or experience with the scenario in a dilemma. Subjects may be more likely to make a DLC when they perceive the positive outcome given doing harm as less likely compared to the one given allowing harm. For instance, in the footbridge vignette, subjects may perceive the probability that the fat man being pushed over the bridge can stop the trolley and thereby five people will be saved being much lower than 100%. The concept of "doing harm" aversion (i.e., prefer a choice of avoiding doing harm to a choice of doing harm) in the footbridge vignette is analogy to the concept of risk aversion (prefer a choice with certainty to a choice with risk) (Rogers, Viding, & Chamorro-Premuzic, 2013).

In a preliminary investigation (Song & Shou, 2014), we used the classical trolley dilemma and footbridge dilemma, and asked subjects for their preference between the CLC and DLC in each dilemma. Depending on their preference, subjects were then asked if they would change their decisions if the outcome probability of their previous decision was not 100%. About 40% subjects, including those who initially chose CLC and those who chose DLC, altered their preferences. In addition, subjects who initially chose CLC were more likely to alter their preference than those who initially chose DLC. It was also found that the proportion of subjects who chose to switch the choices was similar between the trolley dilemma and footbridge dilemma.

Being sensitive to the outcome probability is a substantial feature of a consequentialist reasoning, as it is in accordance with the basic principle of consequentialism---maximizing the expected utility of outcomes (Harsanyi, 1980; Hooker, 2000; Peterson, 2009). The results in Song and Shou (2014) implied that consequentialist reasoning may be applied to generate both CLC and DLC. The equal proportions of subjects who were sensitive to outcome probability across two dilemmas suggest that the proportion of subjects who initiated consequentialist reasoning and subjects who initiated deontological reasoning can be independent from the type of dilemma. Instead, it is a matter of outcome probability.

In the current study, we focused on the impact of outcome probabilities on subjects' moral decisions. We argued that the moral decisions can be influenced by the outcome probabilities, which may offer a new perspective for rethinking the differences in moral decision across different moral dilemma. We used three moral dilemmas that differed in term of the extent to which they are like the personal or the impersonal dilemma. We first assessed subjects' moral judgments without providing any probabilistic information about the outcomes. We hypothesized that like the previous studies, subjects are more likely to choose DLC in a personal dilemma than when presented with an impersonal dilemma.

We then measured subjects' sensitivity to outcome probabilities. Subjects were presented with several paired choices, each of which had the outcome probabilities specified in two choices. If a subject adopts consequentialist reasoning and evaluates the utility of the outcomes, his or her decision should eventually shift to the alternative when the expected outcome utility of their previous decision (the utility of an outcome discounted by its likelihood) becomes lower than the expected outcome utility of its alternative.

In contrast, if a subject evaluates the action itself rather than its outcome, then he or she should be insensitive to the change of probabilities in the outcomes. We hypothesized that subjects who applied consequentialist reasoning could also make the DLC when there was no probability information. That is, there would be subjects who chose DLC initially without probability information shifting their decisions when probability information was provided. Subjects who initially selected CLC were more likely to change the decision than those who initially selected DLC.

In addition, we argue that whether consequentialist or deontological reasoning is applied does not depend on the types of the dilemma. It is a matter of the judgmental model a subject usually adopts. The proportion of subjects who are applying the consequentialist reasoning and are sensitive should be similar across different moral dilemmas. We hypothesized that the proportions of consequentialist reasoning based subjects who eventually shifted their decisions should be independent from the types of moral dilemma.

## Method

### Participants and Procedure

A total of 161 subjects (109 females) were recruited via online crowd-sourcing service CrowdFlower. Subjects aged between 20 and 74, with a mean age of 39.47 years ( $SD = 11.73$ ). Subjects were randomly assigned to one of the three moral dilemmas vignettes described below. They read the consent information, completed the demographical questionnaire, the moral judgment task and the subsequent pairwise choice comparison task in order.

### Materials

There were three moral dilemma vignettes. The detailed vignettes were available in online appendix. The first vignette—*flood* vignette was an impersonal dilemma that was similar to the trolley vignette. The CLC was to initiate the explosion to sacrifice one person to prevent five people from being flooded, while the DLC was to do nothing and allow the five people to be flooded. The decision maker did not have physical contact with the victims and the decisions influence the victims remotely.

The second vignette – *truck* vignette was also an impersonal dilemma. The CLC was about turning the truck into one bystander in order to protect five people in a car, while the DLC was about allowing five people to be hit instead of sacrificing one person. In comparison to the flood case, the truck case had greater distance between the decision maker and the victims in the dilemma.

The final vignette—*hostage* vignette was a personal dilemma and was similar to the footbridge case, where the decision maker needed to physically contact the victim. The CLC was to push a person over the cliff to prevent five hostages from being killed by the gangster, while the DLC was to do nothing and allow five hostages to be killed<sup>1</sup>.

The description of the vignette did not contain any probabilistic information. After reading the vignette, subjects were asked to judge “Which action do you think is morally better?” The question asks subjects to compare two choices in aspects of morality. The question, unlike other common moral judgment questions that ask about “permitted” or “wrong”, can draw subjects’ attention to morality per se from law or convention (Baron, 2014). To engage subjects with the dilemmas, black and white illustrations for each dilemma were presented on screen throughout the whole task.

Next, we specified the probabilities for the outcomes of the two actions and asked subjects to judge which action is more morally correct. For example, the first comparison for the flood dilemma was that “Now suppose we know that the outcomes of your choices may not be 100% sure. Suppose if you choose do nothing, it is 80% sure that the five miners will die. On the other hand, if you choose to explode the floodwall, it is 100% sure that the individual miner will die.

Given this new information, if you are asked to re-do the judgment, which action do you think is morally better?”

We choose to specify the probability of this type of loss (i.e., how likely the victims would die) for the following reasons. First, we avoided the expression that “the victim would be killed” because “kill” implied an action that is not morally neutral, and may bias subjects’ decisions. Second, we avoided using negative wording (i.e., how likely the victims would not die) as subjects may have difficulty in judging the probabilities for the negatively worded statement (Peterson, 2009). Finally, we avoid using vague outcomes such as “how likely the victims would be alive” as the degree of harm relating to being alive is more ambiguous than death.

We changed the probabilities until subjects changed their mind to prefer the alternative choice. Figure 1 illustrates the paired judgments process. The probability-attached choice questions started by a comparison of a DLC with 80% chance of loss against a CLC with 100% chance of loss (the red circle). If a subject chooses CLC, which means that the subject perceives 80% chance of that five people die is morally worse than 100% chance of that one person dies, then he or she proceeds to the second question, in which the probability of the loss in DLC decreased to 20%. If the subject altered the choice and preferred DLC, the third question would increase the probability of the loss in DLC to 50%. After this question, we also narrowed down the possible equivalence of the two choices to the subject. If the subject chose DLC in the third question, then the subject perceives the 100% chance of one death in DLC is morally equivalent to 20% - 50% chance of five deaths in CLC. On the other hand, if the subject chose CLC in the third question, then the subject perceives 100% chance of one death in DLC is morally equivalent to 50% - 80% chance of five deaths in CLC.

Likewise, if a subject chooses DLC in the initial comparison, which means that the subject perceive 100% chance of one death is morally worse than the 80% chance of 5 deaths, then he or she proceeds to the next question, in which the probability of the loss in CLC further decreased. There are eight resultant categories of this decision task as shown in Figure 1, the moral equivalence of the outcomes are summarized in Table 1.

Subjects who stopped in category 1 are those who insisted DLC regardless how low the chance that one individual would die. Stopping in category 8 suggests the subject insisted CLC as the morally better choice even when the expected utility of the two choices are indistinguishable.

<sup>1</sup> The details of materials as well as example illustrations are available in online supplemental materials at <http://goo.gl/hknhMJ>

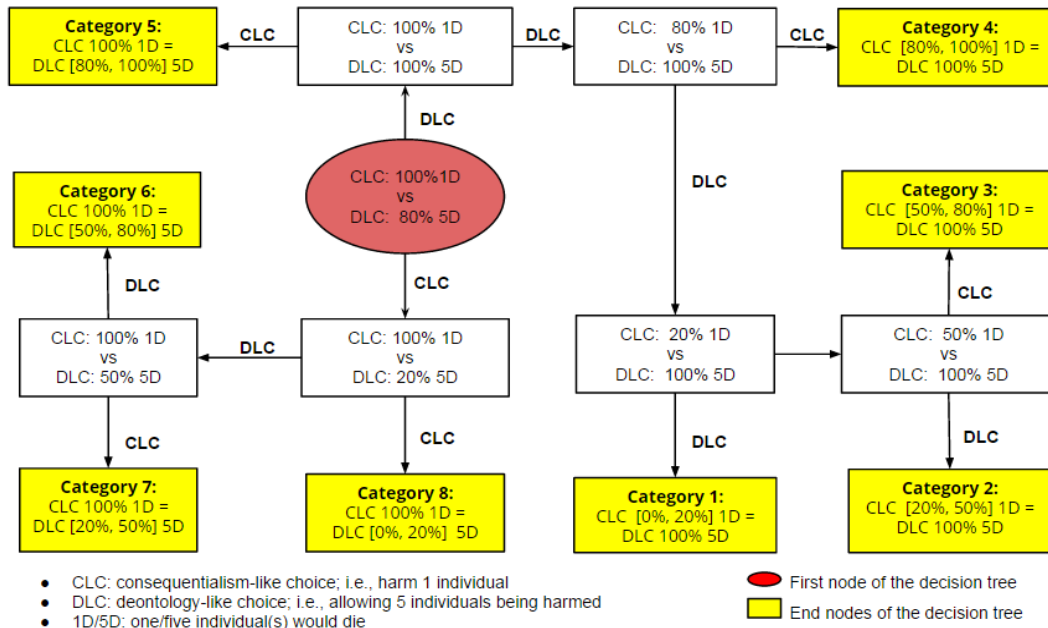


Figure 1. Illustration of the logic flow in probability-attached decision making questions

Table 1: Types of subjects based on the paired comparison where subjects shifted their decisions

Category	Outcome probability	
	DLC (5 deaths)	CLC (1 death)
1	100%	[0%, 20%]
2	100%	[20%, 50%]
3	100%	[50%, 80%]
4	100%	[80%, 100%]
5	[80%, 100%]	100%
6	[50%, 80%]	100%
7	[20%, 50%]	100%
8	[0%, 20%]	100%

## Results

### Moral Decisions

Table 2 displays the frequencies and percentages of the moral decisions across different dilemmas. Logistic regression was conducted to model the choices in different dilemmas by using the CLC as the baseline choice. The type of the moral dilemmas had substantial contribution to the regression model,  $\chi^2 = 33.54$ ,  $p < .001$ , indicating subjects' initial moral decisions were significantly different across different dilemmas. Subjects in the hostage dilemma were slightly less likely to choose CLC than DLC,  $b = -0.51$ ,  $p = .064^2$ . In support of our first hypothesis, subjects in the hostage dilemma were significantly less likely to choose CLC than subjects who were in the other two dilemma,  $b = -1.74$ ,  $p$

$< .001$  in compared to the truck dilemma, and  $b = -2.37$ ,  $p < .001$  in compared to the flood dilemma.

Subjects in the truck and flood dilemma were more likely to choose CLC than DLC,  $b = 1.23$ ,  $p < .001$ , odds ratio = 3.4 for truck dilemma, and  $b = 1.86$ ,  $p < .001$ , odds ratio = 6.4 for flood dilemma. The likelihood of choosing CLC among subjects in the flood dilemma was not significantly different from those in the truck dilemma,  $b = 0.63$ ,  $p = .226$ .

Table 2: Frequencies and percentages of the moral decisions across different dilemmas

Vignette	DLC	CLC	Total
Truck	12 (22.6%)	41 (77.4%)	53
Flood	7 (13.5%)	45 (86.5%)	52
Hostage	35 (62.5%)	21 (37.5%)	56

### Sensitivity to Probability Information

Subjects who were finally in category 1 (insisting DLC) and category 8 (insisting CLC) were regarded as those who were insensitive to probabilistic information. The numbers and proportions of those subjects are displayed in Table 3. On average, the majority of subject (79.2%) were influenced by the change of the probabilities of the outcomes and finally shifted their choices.

<sup>2</sup> We change the dummy coding scheme to obtain the coefficient estimate. The results of the comparisons across the different dilemmas were obtained by conducting three versions of logistic

regression; each of each treats one of the dilemma as the base comparison group.

Table 3: Frequencies and proportion of subjects who did not switch the choice due to the change of probability information

Dilemma	DLC	CLC	Total
Truck	5/12 (41.67%)	3/41 (0.07%)	8/53 (15.1%)
Flood	2/7 (28.57%)	10/45 (22.22%)	12/52 (23.1%)
Hostage	13/35 (37.14%)	0/21 (0%)	13/56 (23.2%)
Total	20/52 (38.46%)	13/107 (12.15%)	38/161 (20.5%)

A logistic regression model was conducted on the likelihood of subjects being sensitive to the different probabilities, predicted by the type of dilemmas and their initial choices. The likelihood of shifting choices was significantly different between subjects who initially chose DLC and those who initially chose CLC,  $\chi^2= 12.98, p <.001$ . As expected by the second hypothesis, subjects who chose CLC were significantly more likely to be influenced by the probabilistic information and changed their choice than those who chose DLC,  $b = 1.84, p <.001$ , odds ratio = 6.2. Furthermore, in support to the third hypothesis, the type of dilemmas did not have significant contribution to the model fit,  $\chi^2= 1.47, p =.479$ , suggesting the proportion of subjects who was sensitive to the probabilities was similar across the three dilemmas.

### Discussion

In the current study, we used three moral dilemmas with similar attributes as the personal/impersonal cases (i.e., trolley case and the footbridge case.) The results were similar to previous studies, where subjects were substantially more likely to choose the consequentialism-like choice (CLC) in the impersonal dilemma (flood and car) than the personal dilemma (hostage). When being provided the probabilistic information about outcomes, about eighty percent of subjects eventually changed their decision on which was a morally better choice. Being influenced by the outcome probability indicated that those subjects might be employing consequentialist reasoning. Subjects who initially chose CLC were more likely to change their choice with the change in probabilistic information, indicating that subjects who chose CLC were more likely to engage in consequentialist reasoning than those who initially chose DLC. This result implied that most people may apply model-based evolution associated with probabilities in moral reasoning. The resultant decisions from the reasoning algorithm are not restricted to a deontology-like decision.

It was also found that the proportions of subjects who were sensitive to the probabilities were similar across the three dilemmas. This suggests that the likelihood of subjects to be engaged in consequentialist reasoning in moral judgments might be independent from the type of moral dilemmas. The large proportion of choice shift in the hostage case among subjects who chose DLC initially, further suggests that the

higher likelihood of choosing DLC in the personal dilemmas can be associated with the perceived probabilities of outcome being different from those in the impersonal dilemmas. As indicated by Crockett (2013) evaluation of the consequences in model-based moral reasoning can be influenced by the prior experience of subjects in associated with the event in the dilemma. The different levels of experience across different events contribute to the different evaluations for the doing harm and allowing harm across different dilemmas.

Interestingly, several subjects who initially chose CLC did not change their choices even when the expected utility of the CLC (the one person has 100% chance to die) was lower than the DLC (the five people have 20% chance to die). One possible explanation is that a subject who chose CLC may also adopt a decision heuristics in system 1, which involve fast and intuitive processes (Evans, 2003). Those subjects made their decisions by comparing the number of loss in each case without evaluating the relative weight of outcomes. We may call them outcome probability insensitive consequentialist reasoner. The other explanation is that, to those subjects, the aggregated utility of five peoples' life can be greater than the single person's life, result in the perceived utility of 20% chance of five people dying as greater than the utility of 100% chance of one person dying.

Another interesting finding was that many subjects changed their decision as soon as the outcome probability of the previous one option had an outcome probability lower than 100%. The majority of subjects stopped at the category where they perceived the consequence of 50% to 80% of the loss (five people would die) in CLC as being equivalent to the 100% of the loss in DLC (one person would die). One possible explanation is that, subjects were risk seeking (i.e., preferred a choice whose outcome is between 0% and 100% over a choice whose outcome is 0% or 100%) in the loss domain even in moral reasoning. Risk seeking behaviors as a result of loss aversion was well documented in decision making literature (d'Acremont et al., 2013). Subjects in the present study might prefer a choice whose probability of loss (five deaths) was lower than 100%, over the alternative choice whose outcome (i.e., one death) probability was 100%.

An alternative explanation is that the outcome utility calculation may also involve the evaluation of action cost. The action of killing may yield action costs associated with social conventions, moral responsibilities and law obligations. That may be the reason why the outcome utility of one death with 100% certainty in CLC is greater than the five deaths with 20% certainty. Both explanations may need further investigation in future studies.

### Limitations and Conclusion

The current experiment demonstrates that most subjects were evaluating consequences in moral reasoning when the probabilistic information was provided. One may argue that, the explicit probability information may induce the adoption of consequentialist reasoning, as consequentialist reasoning, in contrast to deontological reasoning, is a cost-benefit

calculation with probability. This may not be the direct evidence for what people did naturally and intuitively. Future studies in moral dilemmas may assess subjects' prior belief on the probabilities for both the positive outcomes and the negative outcomes to better understand how the probabilistic factor could influence people's moral reasoning.

In conclusion, the results indicate that choosing a DLC does not entail that people engage in consequentialist reasoning, whereas choosing a CLC does not entail that people engage in consequentialist reasoning. The differences across different moral dilemmas are very likely due to the fact that subjects perceive the outcome probability as insufficient for one to choose the alternative choice.

### References

- Baron, J. (2013). *Moral judgment: Acts, omissions, and rules*. Unpublished manuscript, Department of Psychology, University of Pennsylvania, Philadelphia, United States.
- Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, 17(8), 363–6.
- Cummins, D. D., & Cummins, R. C. (2012). Emotion and deliberative reasoning in moral judgment. *Frontiers in Psychology*, 3:328.
- d'Acremont, M., Fornari, E., & Bossaerts, P. (2013). Activity in inferior parietal and medial prefrontal cortex signals the accumulation of evidence in a probability learning task. *PLoS Computational Biology*, 9.
- Evans, J. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7(10), 454–459. doi:10.1016/j.tics.2003.08.012
- Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Penguin Press HC.
- 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 (New York, N.Y.)*, 293, 2105–2108.
- Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6(12), 517–523.
- Harsanyi, J. C. (1980). Rule utilitarianism, rights, obligations and the theory of rational behavior. *Theory and Decision*, 12(2), 115–133.
- Hooker, B. (2000). *Ideal Code, Real World: A Rule-consequentialist Theory of Morality*. Oxford University Press.
- Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. (2014). Emotion and decision making. *Annual Review of Psychology*, 66: 799-823.
- Moll, J., & de Oliveira-Souza, R. (2007). Moral judgments, emotions and the utilitarian brain. *Trends in Cognitive Sciences*, 11, 319–321.
- Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, 36(1), 163–177.
- Peterson, M. (2009). *An Introduction to Decision Theory* (1<sup>st</sup> Ed). New York: Cambridge University Press.
- Rogers, J., Viding, E., & Chamorro-Premuzic, T. (2013). Instrumental and disinhibited financial risk taking: Personality and behavioural correlates. *Personality and Individual Differences*, 55, 645–649.
- Rushworth, M. F. S., & Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience*, 11, 389–397.
- Song, F. and Shou, Y. (2014). Affect and moral judgments, paper presented at *Intentional Conference on Epistemology and Cognitive Science*, Xiamen, China, Jun 2014.
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, 17(6), 476–7.