

Gender Differences in Moral Judgment and the Evaluation of Gender-Specified Moral Agents

Valerio Capraro*, Jonathan Sippel†

May 29, 2017

Forthcoming in Cognitive Processing

*Middlesex University London, United Kingdom. Contact author: V.Capraro@mdx.ac.uk

†University of Amsterdam, The Netherlands.

Abstract

Whether, and if so, how exactly gender differences are manifested in moral judgment has recently been at the center of much research on moral decision making. Previous research suggests that women are more deontological than men in personal, but not impersonal, moral dilemmas. However, typical personal and impersonal moral dilemmas differ along two dimensions: personal dilemmas are more emotionally salient than impersonal ones *and* involve a violation of Kant's practical imperative that humans must never be used as a mere means, but only as ends. Thus, it remains unclear whether the reported gender difference is due to emotional salience or to the violation of the practical imperative. To answer this question, we explore gender differences in three moral dilemmas: a typical personal dilemma, a typical impersonal dilemma, and an intermediate dilemma, which is not as emotionally salient as typical personal moral dilemmas, but contains an equally strong violation of Kant's practical imperative. While we replicate the result that women tend to embrace deontological ethics more than men in personal, but not impersonal, dilemmas, we find no gender differences in the intermediate situation. This suggests that gender differences in these type of dilemmas are driven by emotional salience, and not by the violation of the practical imperative. Additionally, we also explore whether people think that women should behave differently than men in these dilemmas. Across all three dilemmas, we find no statistically significant differences about how people think men and women should behave.

Keywords: gender differences, morality, deontology, consequentialism, practical imperative.

1 Introduction

Whether, and if so, how exactly gender differences are manifested in moral judgment has recently been at the center of much research on moral decision making. This research is usually traced back to Kohlberg's six stage theory of moral reasoning (e.g. Kohlberg (1969)) and Gilligan's opposing work, which, amongst other things, criticised Kohlberg's theory as being biased in favor of men (Gilligan, 1982). Kohlberg stated that the moral development, more precisely the development of one's justifications of one's moral actions, occurs on six stages, where reaching a new stage makes one more adequate at responding to moral dilemmas than the ones before. In his theory, moral development however is mainly concerned with attaining a sense of *justice* and *fairness*. According to Gilligan, Kohlberg perceived women to usually get stuck at level three of this development, while men quite often move forward to more abstract principles of morality. She further notices that Kohlberg's first study included only adolescent boys (Gilligan, 1982). Furthermore, implicit in Kohlberg's theory is the exclusion of other aspects of moral behaviour than justice and fairness. Gilligan countered this claim by proposing a theory of moral development that was not focused on justice but on *caring* (focusing on maintaining relationships, responding to the needs of others and not to cause any harm). This move was supposed to account better for the moral development of females.

Since then, a lot of empirical work focusing on an investigation of actual gender differences in moral decision making has been done - with diverging results. While some found a statistically significant difference in moral decision making between women and men (Björklund, 2003; Aldrich & Kage, 2003; Gump, Baker, & Roll, 2000; Zamzow & Nichols, 2009; Banerjee, Huebner, & Hauser, 2010), others did not (Jaffee & Hyde, 2000; Kaesey, 1972; Seyedsayamdost, 2015).

Kohlberg's and Gilligan's theories provide the ground for some explanatory accounts of any supposed gender-difference in moral judgment: Björklund (2003, 459), inspired by the debate between Kohlberg and Gilligan, distinguishes between care-oriented and justice-oriented moralities, where the former and the latter are supposedly more present in women and men, respectively. In a study, Björklund shows that making the difference between ethics of care and ethics of justice

indeed introduces gender differences, arguing that moral judgment criteria vary due to socialisation - often leaving males and females with differing criteria to evaluate moral dilemmas. However, other studies suggested that there is no strong empirical support for the assertion that women and men predominantly make use of the care orientation and the justice orientation, respectively (Jaffee & Hyde, 2000).

In this work, we aim to contribute to this ongoing debate about when exactly these supposed gender differences appear. For our study, we will differentiate three versions of a moral dilemma (the well-known Trolley Problem that is introduced below) in which participants have to choose between two courses of action - one being characteristically consequentialist, the other being characteristically deontological. Consequentialist ethics value actions according to the consequences they bring about and are often said to have their root in the utilitarianism of Jeremy Bentham and John Stuart Mill: “actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness” (Mill, 1863/1906, 9). Deontological ethics employ categorical principles that guide the evaluation of the moral value of an action. A crucial example that will play a role in this investigation is Immanuel Kant’s *practical imperative* that demands respect for human beings as such: “So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means” (Kant, 1785/1998, 4:429). Deontological ethics are often said to best capture moral intuitions that people actually have. The different choices in moral dilemmas will henceforth be discussed within this terminology.

Following the work of Greene *et al.* (2001; 2004), we will further distinguish moral dilemmas in two subclasses, personal and impersonal dilemmas. Greene *et al.* (2001) linked an apparent difference concerning moral judgment in the cases of impersonal and personal moral dilemmas to the fact that participants are more emotionally involved in the latter case than in the former. Using fMRI scans, they show that their subjects’ brain areas associated with emotion prove to be significantly more active when contemplating a personal dilemma than a non-personal one, thus motivating a qualitative distinction of moral dilemmas in terms of personal emotional involvement. As result of their study, a proto-typical impersonal moral dilemma is the standard Trolley Problem, while a proto-typical personal moral dilemma is the Footbridge Dilemma (we refer to next section for

the exact formulations of these dilemmas). This distinction turns out to be relevant in this context because a recent study by Fumagalli *et al.* (2010) found that men tend to embrace consequentialist judgment significantly more than women, but only in the case of *personal moral dilemmas*.

The reason for this gender difference however remains unclear, because, as it will be detailed in the next section, the personal moral dilemmas adopted by Fumagalli *et al.* (2010) differ from the impersonal ones in two dimensions: not only their personal moral dilemmas are more emotionally salient than their impersonal ones, but they also involve a strong violation of the practical imperative. The *practical imperative* introduced above prohibits to use other people as a mere means. It is thus possible that the gender difference between the evaluation of personal and impersonal moral dilemmas of the sort that Fumagalli *et al.* (2010) introduced was not due to the fact that personal moral dilemmas are more emotionally salient, but simply to the fact that the consequentialist course of action in them represents a massive violation of the practical imperative not to treat others as a mere means. Their study thus leaves it unclear whether the reported gender difference is due to emotional salience or the violation of the practical imperative. To differentiate between these two possibilities, we will therefore introduce a moral dilemma further below that is somewhat *in the middle* of the personal and impersonal moral dilemmas that were investigated by Fumagalli *et al.* (2010): it contains an almost equally strong violation of the practical imperative, but is much less emotionally salient than the one investigated by Fumagalli *et al.* (2010).

Furthermore, we wish to add another perspective on the role of gender in moral judgment: we will change the moral dilemmas such that participants either evaluate a man's (Adam) or a woman's (Amanda) moral choices. By this we wish to not only investigate whether there are gender differences in the evaluation of moral dilemmas but also its mirror image, i.e. whether there are differences in the evaluation of gender-specified moral agents (this gender specification is done by means of naming the moral agents in our dilemmas either Amanda or Adam). We found the existing evidence on this question to be quite unsatisfying - in comparison to the question whether gender differences are manifested in moral judgment, relatively few research has been done. A meta-analysis by Jaffee and Hyde (2000) found that the resulting evidence is at best inconsistent (Bussey & Maughan, 1982; Freeman & Giebink, 1979; Garwood, Levine, & Ewing, 1980; Krebs,

Vermeulen, Denton, & Carpendale, 1994; Lonky, Roodin, & Rybash, 1988; Orchowsky & Jenkins, 1979; Turiel, 1976): Discussing the role of the gender of the protagonist in a moral dilemma in terms of the justice-care dichotomy, Albrecht (1989) found that dilemmas with male protagonists lead to significantly more reasoning according to the justice-orientation than dilemmas with female protagonists. Bussey and Maughan (1982) found a significant difference in the evaluation of gender-specified moral agents by men (evaluating their actions according to the care-oriented morality in case the agent was specified as female and according to the justice-oriented morality otherwise) - but not in the case of female subjects. Garwood *et al.* (1980) on the other hand found no such evidence at all. There is thus no broad consensus on this question in the literature - which is not surprising as the results tend to vary with experimental design choices. Furthermore, the literature does, to the best of our knowledge, not provide any theories to ground a strong hypothesis regarding these issues. Given the inconsistent evidence that nonetheless tends towards refuting any (general) significant effect of specifying the gender of a moral agent, we will thus hypothesise that such an effect does not exist. As there are presumably gender differences in moral judgment, one may wonder (and test) whether observers think that women should behave differently than men.¹

This paper is structured as follows. Next section contains a study overview, which includes all the relevant definition. Then we investigate whether the evaluation of third-person moral dilemmas varies when the gender of the agent in these dilemmas is specified as female or male. We further revisit the results obtained by Fumagalli *et al.* (2010), by investigating whether analogous gender differences can also be found in another moral dilemma that is somewhat *in between* theirs. Finally, we speculate on the relevance and the limitations of our results.

¹One further remark is in order here: Kantian deontological ethics would not allow to treat different genders differently. Kantians might thus not be very happy about the fact that we investigate such behavior in Kantian terms. Moral agents can however not be expected to be fully consistent – and our results will indeed prove that they will uphold Kantian principles in certain situations but not in others. The lurking contradiction is thus only an apparent one that arises from making an empirical, descriptive study using terminology from a normative theory: we do not claim that our participants are either (Kantian) deontologists or not, but that they may or may not pick the characteristically deontological course of action, i.e. upholding the practical imperative, in such and such a situation.

2 Study Overview

We aim to investigate the impact of gender on the evaluation and choice of the characteristically deontologist and consequentialist courses of action in moral dilemmas. This is achieved by presenting subjects an *external* version of moral dilemmas in which they evaluate another person's possible choices.

To investigate which principles give rise to gender differences, participants are randomly assigned to make a choice in one of three moral dilemmas (introduced below). Furthermore, to investigate whether decision makers of different gender are expected to behave differently in moral dilemmas, participants are randomly assigned to one of two conditions in which the decision maker is either a man (Adam) or a woman (Amanda). This corresponds in sum to *six* different conditions to exactly one of which the participants are randomly assigned. We will now shortly introduce the three moral dilemmas in the case of Amanda (the case of Adam works analogous) and then offer some important considerations about our dilemmas.

Trolley Problem A runaway trolley is heading down the tracks toward five workers who will all be killed if the trolley proceeds on its present course. Amanda is standing next to a large switch that can divert the trolley onto a different track. The only way to save the lives of the five workers is to divert the trolley onto another track that only has one worker on it. If Amanda diverts the trolley onto the other track, this one worker will die, but the other five workers will be saved.

Trapdoor Dilemma A runaway trolley is heading down the tracks toward five workers who will all be killed if the trolley proceeds on its present course. Amanda is on a footbridge over the tracks, in between the approaching trolley and the five workers. Next to her on this footbridge is a stranger who happens to be very large. The only way to save the lives of the five workers is to flip a switch to release a trapdoor that will drop the stranger off the bridge and onto the tracks below where his large body will stop this trolley. The stranger will die if Amanda does this but the five workers will be saved.

Footbridge Dilemma A runaway trolley is heading down the tracks toward five workers who will all be killed if the trolley proceeds on its present course. Amanda is on a footbridge over the tracks, in between the approaching trolley and the five workers. Next to her on this footbridge is a stranger who happens to be very large. The only way to save the lives of the five workers is to push the stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die if Amanda does this, but the five workers will be saved.

Moral dilemmas of this kind were first introduced by Philippa Foot (1967). They provide useful insights in the subjects' behaviour because they are artificial, moral situations of very distinct and specific quality. They not only have different wording but also come with strong ethical implications. Given that we know in what way the three moral dilemmas above differ, we can trace back gender differences manifested in their evaluation to the specific and distinct features of the dilemmas. Fumagalli *et al.* (2010)'s research on the Trolley Problem and the Footbridge Dilemma led them to assume that this difference is caused by the fact that the latter is much more emotionally salient. But they also differ in their violation of the practical imperative: in the Footbridge Dilemma the death of the stranger is an explicit part of Amanda's plan; whereas in the Trolley Problem this death is merely a foreseeable consequence of this plan - but not an essential part of it. Choosing the consequentialist option in the Trapdoor Dilemma is thus a stronger violation of Kant's practical imperative. This is why we introduce the Trapdoor Dilemma to our investigation: it involves a violation of the practical imperative as strong as the Footbridge Dilemma, but it is not as emotionally salient (Everett, Pizarro, & Crockett, 2016; Capraro *et al.*, 2017). These three dilemmas were thus picked firstly to force participants to decide between two courses of action, one of which is characteristically deontological, while the other is characteristically consequentialist, and secondly to give them different shades of involved violence or emotional involvement. This accounts for the fact that being a deontologist (or consequentialist) is rarely a yes or no question. By this, we aim to obtain a clearer picture of what causes the gender differences observed by Fumagalli *et al.* (2010).

We will thus investigate in the first part of our study whether our subjects evaluate the moral dilemma of a gender-specified moral agent differently in the cases of Amanda and Adam. In our

scenario that builds on third-person moral dilemmas, it is highly important to do this before the second part of the study. In this second part of the study, we will evaluate our results to see whether gender differences of our subjects are manifested in their moral judgment.

3 Evaluation of Gender-Specified Moral Agents

In the first part, we wish to investigate whether people evaluate the three moral dilemmas above differently in cases where the agent in the dilemmas is a woman or a man.

Hypothesis The evaluation of the gender-specified moral agents Amanda and Adam does not differ. That is, there exists no statistically significant difference in the evaluation of the respective Amanda and Adam cases of our three dilemmas.

Method *Participants.* 600 American participants were recruited using *Amazon Mechanical Turk*.² They were paid \$0.40 for their time. A total of 68 subjects were excluded because they either failed the comprehension questions or took the survey more than once, leaving us with a final sample of 532 participants.

Design. Participants were randomly assigned to one of the three moral dilemmas described above (Trolley Problem, Trapdoor Dilemma, Footbridge Dilemma). Furthermore, they were then randomly assigned to report whether either Amanda or Adam should take the consequentialist course of action and thereby kill the stranger.

Results and Discussion In accordance with the hypothesis, there was no significant difference in the evaluations of the *Amanda* and *Adam* tasks (see figure 1). In the Trolley Problem, 32% chose the deontological option for Amanda, 24% for Adam (leaving 68% and 76% consequentialists, respectively). In the Trapdoor Dilemma, 52% responded that Amanda should take the deontological

²The reliability of *Amazon Mechanical Turk* has been thoroughly investigated (e.g. (Paolacci, Chandler, & Ipeirotis, 2010; Horton, Rand, & Zeckhauser, 2011; Mason & Suri, 2012; Brañas-Garza, Capraro, & Rascón-Ramírez, 2016; d’Adda, Capraro, & Tavoni, 2017)).

course of action, while 55% voted the same for Adam (leaving 48% and 45% consequentialists, respectively). As Fumagalli *et al.* (2010) noted a gender difference in the evaluation of the Footbridge Dilemma, we thought that at least in this case, Amanda might be evaluated differently than Adam. However, only 59% responded that Amanda should take the deontological course of action while 64% responded the same for Adam (leaving 41% and 36% consequentialists, respectively). None of the three cases does the fact whether the subjects were confronted with either the *Amanda* or *Adam* task significantly correlate with choosing the deontological or consequentialist course of actions (all p-values > 0.20).³

Conclusion These results show that women's and men's actions in our three moral dilemmas are not evaluated differently when the gender is only indicated by the names *Amanda* and *Adam*.

4 Gender Differences in Moral Judgment

In the second part, we test the results provided by Fumagalli *et al.* (2010) that there is a gender difference in moral judgment. Furthermore, we wish to investigate further how these results come about. Fumagalli *et al.* (2010) use moral dilemmas resembling our Trolley Problem and Footbridge Dilemma.⁴ This distinction however does not make clear the cause of the difference in the evaluation of the Footbridge Dilemma, as there are two fundamental differences between the Trolley Problem and the Footbridge Dilemma: the consequentialist option in the Footbridge Dilemma does not only contain a straightforward violation of Kant's practical imperative (not to use people as

³A logit regression without control on gender, age and education in the Trolley Problem led to p-value 0.2775, χ^2 (1) = 6.39913, coeff=0.353995, z=1.086. With control, we obtained p-value 0.2688, χ^2 (4) = 6.99903, coeff=0.364350, z=1.106. A logit regression without control on gender, age and education in the Trapdoor Dilemma led to p-value 0.6446, χ^2 (1) = 0.331016, coeff=-0.141944, z=-0.4612. With control, we obtained p-value 0.6668, χ^2 (4) = 14.6967, coeff=-0.138823, z=-0.4305. Finally, a logit regression without control on gender, age and education in the Footbridge Dilemma led to p-value 0.4718, χ^2 (1) = 1.75243, coeff=-0.225039, z=-0.7195. With control, we obtained p-value 0.6433, χ^2 (4) = 10.4933, coeff=-0.151055, z=-0.4631.

⁴In fact, the work of Greene *et al.* (2001, 2004) that they rely on treats the Trolley Problem and the Trapdoor Dilemma as being characteristic of impersonal and personal moral dilemmas, respectively.

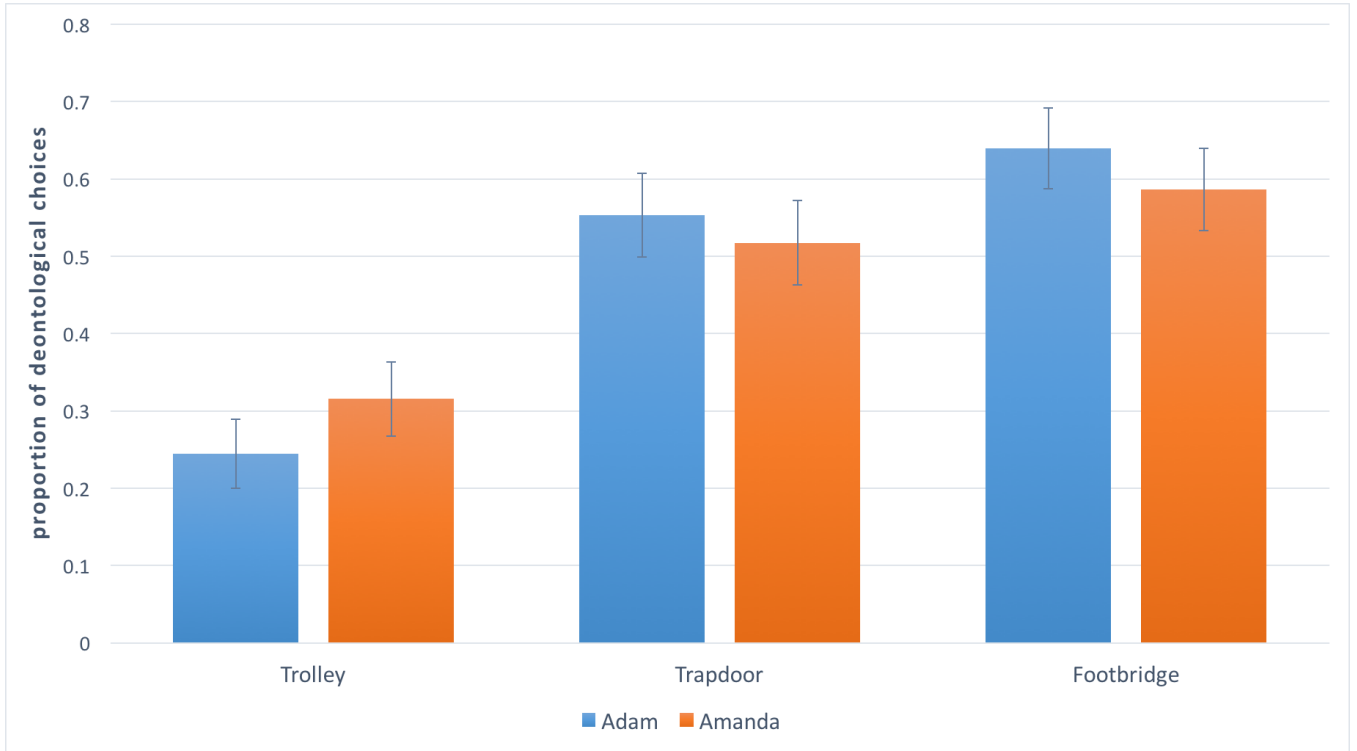


Figure 1: Proportion of people that take the deontological course of action in all three moral dilemmas. Error bars represent the standard error of the mean. There is no significant difference in whether people are confronted with an *Amanda*- or *Adam*-task.

mere means), but also the much more violent, *emotionally involving* act of pushing a stranger off a bridge and towards death with one's bare hands. To get a more fine-grained picture of these gender differences, we treat the Trapdoor Dilemma as an intermediate case that contains a use of another person as a mere means but not a blatantly violent, emotionally involving act like the Footbridge Dilemma. Thus, if we can reproduce Fumagalli *et al.*'s (2010) findings, the additional evaluation of the Trapdoor Dilemma leads to a clearer view of how these findings come about - i.e. whether they are connected to using people as a mere means or emotionally involving violence.

Hypothesis There are no gender differences in the evaluation of the three moral dilemmas introduced as above.

Method We made use of the same data as in Part 1. Participants and design are thus the same as there.

Results and Discussion Results are summarized in Figure 2. In line with Fumagalli *et al.* (2010), there was no statistically relevant difference of gender in moral judgement in the Trolley Problem: 27% of the females and 30% of the males chose the deontological course of action, respectively. But as expected, this result fails to be significant (p-value 0.6775).⁵

In the appended intermediary Trapdoor Dilemma, 54% of the females and 53% of the males chose the deontological course of action, leaving 46% and 47% consequentialists, respectively. As before, this result fails to be significant (p-value 0.8435).⁶

In the Footbridge Dilemma however, 71% of the females and 53% of the males chose the deontological course of action, leaving 29% and 47% consequentialists, respectively. This difference turns out to be statistically significant (p-value 0.0161).⁷ Hence, there is a significant correlation of gender and choosing the deontological course of action in the Footbridge Dilemma, but neither in the Trapdoor Dilemma nor the Trolley Problem.

Conclusion In light of these results, we can reject the hypothesis that there are no gender differences manifested in moral judgment. As Fumagalli *et al.* (2010), we found a statistically significant difference in the Footbridge Dilemma, but not in the Trolley Problem. Additionally, we found that there are no gender differences manifested in the evaluation of the intermediate Trapdoor Dilemma. This leads to the conclusion that it is *not* the violation of the practical imperative but its emotional salience that leads to the difference observed in the Footbridge Dilemma.

⁵A logit regression without control on age, education and whether subjects were assigned to either the Amanda or Adam task led to $p=0.6775$, $\chi^2(1) = 5.38597$, $\text{coeff}=-0.136336$ and $z=-0.4159$. With control, we obtained $p=0.6953$, $\chi^2(4) = 6.99903$, $\text{coeff}=-0.132789$ and $z=-0.3916$.

⁶A logit regression without control on age, education and whether subjects were assigned to either the Amanda or Adam task led to $p=0.8435$, $\chi^2(1) = 0.157122$, $\text{coeff}=0.0607937$ and $z=0.1974$. With control, we obtained $p=0.7222$, $\chi^2(4) = 14.6967$, $\text{coeff}=-0.117436$ and $z=-0.3555$.

⁷A logit regression without control on age, education and whether subjects were assigned to either the Amanda or Adam task led to $p=0.0161$, $\chi^2(1) = 7.17284$, $\text{coeff}=0.772383$ and $z=2.407$. With control, we obtained $p=0.0443$, $\chi^2(4) = 10.4933$, $\text{coeff}=0.659502$ and $z=2.011$.

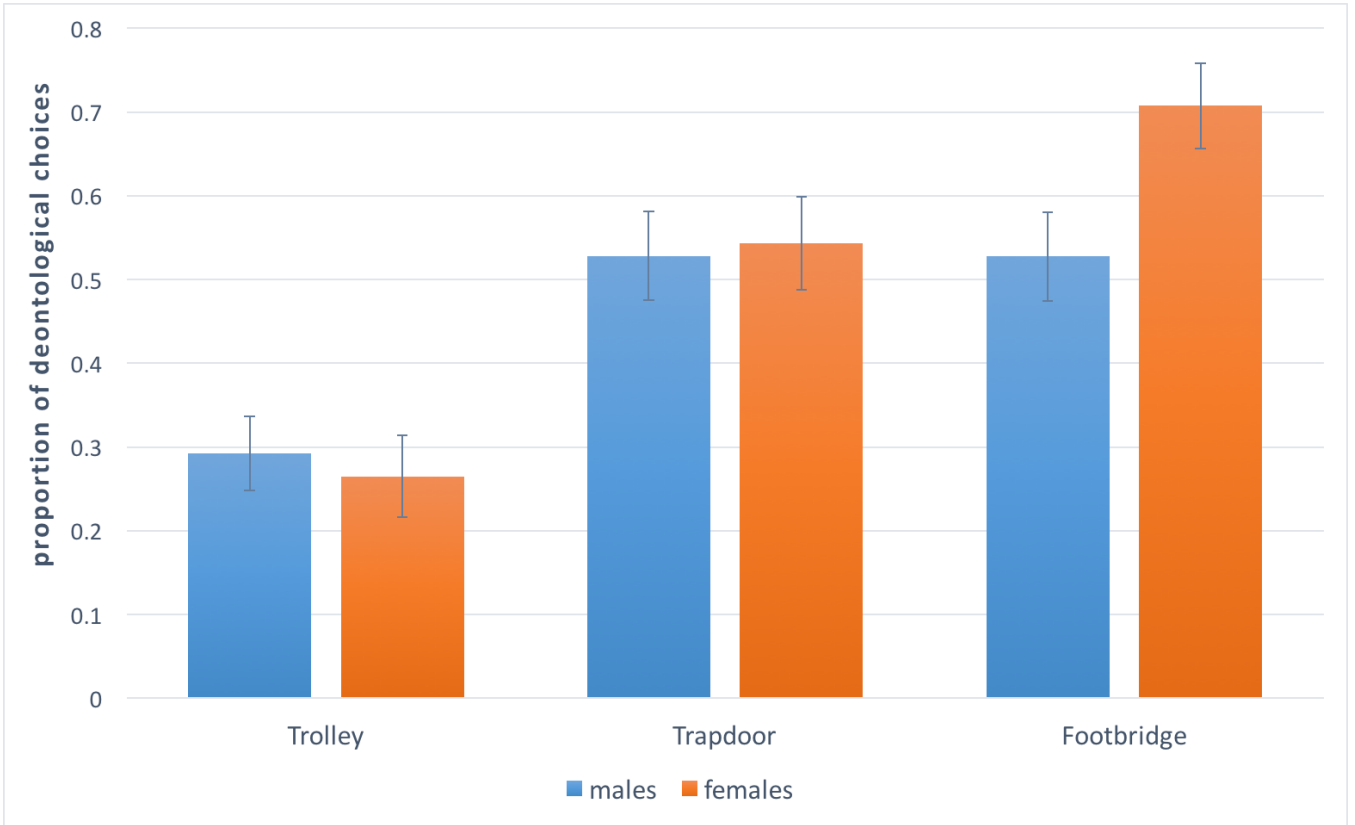


Figure 2: Proportion of people that take the deontological course of action in all three dilemmas. Error bars represent the standard error of the mean. There is a significant difference between females and males in the Footbridge Dilemma but not in the others. The study thus confirms Fumagalli *et al.*'s findings while adding the additional information that the gender-bias in the evaluation of these dilemmas does not come about due to the violation of the classical deontological imperative not to use others as mere means, but is due to the high emotional salience of the consequentialist course of action in the Footbridge Dilemma.

5 Conclusions and Discussion

The first part of our study investigated whether the gender of moral agents matters in the evaluation of their judgments by other people. Our evidence suggests that this is *not* the case. However, some remarks are in order. It should be noted that the difference between the Amanda and Adam tasks was simply the substitution of one name with another. Subjects that are confronted with the task might be fully absorbed by the understanding of their options - too absorbed to notice the gender of the moral agent. An option for future experiments might be to give people a stronger trigger than

the names Amanda and Adam - e.g. by using pictures. As a meta-study (Jaffee & Hyde, 2000) found at best inconsistent evidence on this question, we are convinced that the question of how exactly the form of a moral dilemma influences decision making is of great importance. Without paying attention to this side of the question, a too one-sided focus on how gender, religion and other factors might influence decision making can lead to misleading results. The first part of our study however allows us to exclude this possibility from our investigation.

The second part of our study confirmed and further investigated the results of Fumagalli *et al.* (2010). Their findings and ours suggest that the question whether there are any consistent gender differences in moral judgment does not take a yes or a no for an answer - the situation is more complicated. As mentioned in the introduction, some researchers have claimed to have found gender differences in moral judgment, while others did not. The second part of our study not only adds another piece of evidence to this research but tries to answer another question: are there gender differences in moral judgment, and if so, *when* do they occur?

The use of moral dilemmas of distinct qualities allows us to narrow down this *when*-question. It allows us to say that, if a dilemma has such and such clearly distinguished qualities (as the Footbridge Dilemma), gender differences occur. And if it different but still clearly distinguished qualities (as in the Trolley Problem and the Trapdoor Dilemma), they do not occur. We can now make reasonable assumptions on what quality of the Footbridge Dilemma accounts for the observed difference. As already mentioned in the discussion of the second part of our study, this difference can not be attributed to the violation of Kant's practical imperative - had this been the case, we would have found a significant difference in the Trapdoor Dilemma as well. Using the terminology brought forward by Greene *et al.* (2001), this difference can now really be traced back to the personal, emotional involvement brought about by the brutality of the Footbridge Dilemma.

References

- Albrecht, E. F. (1989). The relationship between protagonist gender in hypothetical moral dilemmas and the use of justice versus care moral orientation (doctoral dissertation, adelphi university, 1989. *Dissertation Abstracts International*, 49, 1734A.
- Aldrich, D., & Kage, R. (2003). Mars and venus at twilight: A critical investigation of moralism, age effects, and sex differences. *Political Psychology*, 24, 23–40.
- Banerjee, K., Huebner, B., & Hauser, M. (2010). Intuitive moral judgments are robust across variation in gender, education, politics and religion: A large-scale web-based study. *Journal of Cognition and Culture*, 10(3), 253–281.
- Björklund, F. (2003). Differences in the justification of choices in moral dilemmas: Effects of gender, time pressure and dilemma seriousness. *Scandinavian Journal of Psychology*, 44, 459–466.
- Brañas-Garza, P., Capraro, V., & Rascón-Ramírez, E. (2016). Gender differences in altruism: Expectations, actual behaviour and accuracy of beliefs. *Available at SSRN: <https://ssrn.com/abstract=2796221>*.
- Bussey, K., & Maughan, B. (1982). Gender differences in moral reasoning. *Journal of Personality and Social Psychology*, 42, 701–706.
- Capraro, V., Sippel, J., Zhao, B., Hornischer, L., Savary, M., Terzopoulou, Z., ... Griffioen, S. F. (2017). Are kantians better social partners? people making deontological judgments are perceived to be more prosocial than they actually are. *Available at SSRN: <https://ssrn.com/abstract=2905673>*.
- d’Adda, G., Capraro, V., & Tavoni, M. (2017). Push, don’t nudge: Behavioral spillovers and policy instruments. *Economics Letters*, 154, 92-95.
- Everett, J. A. C., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, 145, 772–787.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5,

5–15.

- Freeman, S. J., & Giebink, J. W. (1979). Moral judgment as a function of age, sex, and stimulus. *Journal of Psychology, 102*, 43–47.
- Fumagalli, M., Ferrucci, R., Mameli, F., Marceglia, S., Mrakic-Sposta, S., Zago, S., . . . Priori, A. (2010). Gender-related differences in moral judgments. *Cognitive Processing, 11*(3), 219–226.
- Garwood, S. G., Levine, D. W., & Ewing, L. (1980). Effect of protagonist's sex on assessing gender differences in moral reasoning. developmental psychology. *Developmental Psychology, 16*, 677–678.
- Gilligan, C. (1982). *In a different voice*. Harvard University Press.
- Greene, J. D., Nystrom, L. E., Darley, J. M., Engell, A. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389–400.
- 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.
- Gump, L. S., Baker, R. C., & Roll, S. (2000). Cultural and gender differences in moral judgment: a study of mexican americans and anglo-americans. *Adolescence, 35*(137), 67–76.
- Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics, 14*(3), 399–425.
- Jaffee, S. R., & Hyde, J. S. (2000). Gender difference in moral orientation: A meta-analysis. *Psychological Bulletin, 126*(5), 703–726.
- Kaesity, C. B. (1972). The lack of sex differences in the moral judgments of preadolescents. *The Journal of Social Psychology, 86*(1), 157–158.
- Kant, I. (1785/1998). *Groundwork of the metaphysics of morals*. Cambridge: Cambridge University Press.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. *The handbook of socialization theory and research*, 347–489.
- Krebs, D. L., Vermeulen, S. C., Denton, K. L., & Carpendale, J. I. (1994). Gender and perspective differences in moral judgment and moral orientation. *Journal of Moral Education, 23*, 17–26.
- Lonky, E., Roodin, P. A., & Rybash, J. M. (1988). Moral judgment and sex role orientation as a

- function of self and other presentation mode. *Journal of Youth and Adolescence*, *17*, 189–195.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon’s Mechanical Turk. *Behavior Research Methods*, *44*(1), 1.
- Mill, J. S. (1863/1906). *Utilitarianism*. Chicago: The University of Chicago Press.
- Orchowsky, S. J., & Jenkins, L. R. (1979). Sex biases in the measurement of moral judgment. *Psychological Reports*, *44*, 1040.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411-419.
- Seyedsayamdost, H. (2015). On gender and philosophical intuition: Failure of replication and other negative results. *Philosophical Psychology*, *28*(5), 642–673.
- Turiel, E. (1976). A comparative analysis of moral knowledge and moral judgment in males and females. *Journal of Personality*, *44*, 195–208.
- Zamzow, J. L., & Nichols, S. (2009). Variations in ethical intuitions. *Philosophical Issues*, *19*(1), 368–388.