Sacrificing oneself or another: The difference between prescriptive and normative judgments in moral evaluation

Štěpán Bahník^{1*}, Emir Efendic², and Marek A. Vranka¹

¹ Faculty of Business Administration, University of Economics, Prague, náměstí Winstona Churchilla 4, Prague, 130 67, Czech Republic.

² Laboratoire de Psychologie Santé et Qualité de Vie (EA4139), Université de Bordeaux, 3ter place de la Victoire, 33076, Bordeaux Cedex, France, emir.efendic@u-bordeaux.fr

* Corresponding author, bahniks@seznam.cz

Abstract

In situations where one has the opportunity to either sacrifice oneself or another person to save a group of people, one might think that there are instances where people would rather sacrifice themselves than others – thus showing an other-serving bias. So far however, most studies found instances of a self-serving bias – people would rather sacrifice others. In this paper, we test whether an other-serving bias might appear as a function of judgment type. That is, we looked at prescriptive (should/should not) and normative (wrong/right) types of judgments separately. We found that participants exhibited an other-serving bias, but only when asked whether self- or other-sacrifice is wrong. Otherwise, participants were exhibiting a self-serving bias; that is, they approved sacrificing others more. The results underscore the importance of question wording and suggest that some effects on moral judgment might depend on the type of judgment.

Keywords: moral judgment, self-other asymmetry, trolley dilemma, self-sacrifice

People are known to exhibit various self-other asymmetries in their judgments. These asymmetries usually manifest in the form of a self-serving bias. For instance, people see positive attributes as being more descriptive of themselves than of others (Brown, 1986), they expect others to be more affected by alcohol than themselves (Rohsenow, 1983), and they see many more operational, as well as cognitive, biases in others than in themselves (Pronin, Lin, & Ross, 2002). Additionally, people are also over-optimistic about positive, but not negative, future events when comparing themselves with others (Alloy & Ahrens, 1987; Hoorens, 1995). Similar asymmetries have also been observed in moral judgment. For example, in the case of moral hypocrisy, people judge moral transgressions, committed by themselves, less harshly than the

same acts committed by others (Valdesolo & DeSteno, 2007; 2008), and they consider themselves to be in general more ethical, charitable, cooperative, and fair than others (Epley & Dunning, 2000).

However, the self-other asymmetries are not always necessarily self-serving. In some cases people exhibit other-serving biases – their judgments are biased in favor of others, rather than themselves. For example, when participants could have paid for avoiding electrical shocks, they were more willing to pay for others than for themselves (Crockett, Kurth-Nelson, Siegel, Dayan, & Dolan, 2014).

Similarly, one would expect that in studies using trolley (and similar) dilemmas¹, people ought to consider self-sacrifice (i.e., and other-serving bias) as the morally right action. Indeed, anecdotal evidence (Huebner & Hauser, 2011), as well as some research findings (Lester, 2014) suggest that many participants do in fact think this. Although self- and other-sacrifice are supposed to be equivalent from a purely consequentialist point of view, some philosophers argue that the overall amount of harm or good done is not the only thing that matters when such actions are evaluated (Slote, 1984; Stocker, 1976). Discounting one's own pleasures and pains, and therefore exhibiting other-serving bias, seems to be morally permissible and even praiseworthy (Splawn, 2001). For example, although the action of stabbing causes harm and can therefore be considered bad, it seems difficult to judge stabbing oneself as being similarly wrong as stabbing someone else. Similarly, in the case of the trolley dilemma, it is difficult to judge sacrificing oneself to be as wrong as sacrificing someone else.

Consequently, while we would expect an other-serving bias in the evaluation of sacrifice for the sake of other people, Huebner and Hauser (2011) failed to conclusively demonstrate this in their study. In their modified trolley scenario, participants had to decide which of the following three actions they should do: turning a trolley to a track with an anonymous person, letting the trolley run over five people by doing nothing, and turning the trolley to a track where they were standing themselves. While one third of participants answered that they should sacrifice themselves, almost half of the participants said that they should save the five people by turning the trolley to the anonymous person. The results thus suggest a self-serving bias. However, a limitation of the study lies in the joint evaluation of all three possible actions in the trilemma scenario. Some people may be indifferent between both possible sacrifices and choose one of them randomly or decide not to choose any. It is therefore difficult to infer the extent of the self- or other-serving bias in moral judgment using this type of a trilemma scenario.

Notwithstanding this methodological issue, the type of judgment people were asked to make may help explain why the self-serving bias was observed instead of the expected other-serving bias. Specifically, participants in the Huebner and Hauser's study made prescriptive moral

¹ These kind of dilemmas, named after the most famous example by Foot (1978), present a scenario in which a larger number of people can be saved by sacrificing another person. A common interpretation is that two moral principles are pitted against each other in these dilemmas: from the "utilitarian" perspective it is right to save more people while from the "deontological" perspective it is wrong to knowingly cause harm by sacrificing the person. Thus it is wrong to sacrifice anyone even in order to save more people (Greene, 2007).

judgments² about what they *should do*, rather than normative judgments about what is *right* or *wrong* to do. However, it was shown that prescriptive and normative judgments of trolley-like dilemmas differ in significant ways. For instance, people were more condemning of a utilitarian action when asked for prescriptive judgment of what should be done, than when they make normative judgments of whether the action is wrong (Barbosa & Jiménez-Leal, 2017). It is also possible that these different judgment types are influenced by different aspects of judged dilemmas. For example, when making a normative judgment (i.e., “is it wrong/right to do it”), one may more likely focus solely on the morality of the action itself. On the other hand, when making a prescriptive judgment (i.e., “should I do it?”), one may take a wider context into account and consider also what consequences the action will bring about for oneself. It was, for example, showed that utilitarians are perceived as cold, calculating, and untrustworthy (Everett, Pizarro, & Crockett, 2016; Uhlmann, Zhu, & Tannenbaum, 2013). Because people may want to avoid being perceived as possessing such negative characteristics, they refrain from expressing support for the utilitarian action (Rom & Conway, 2018).

That different types of judgment may lead to different moral evaluations was observed in previous studies which showed lower approval of utilitarian actions in case of normative, rather than descriptive judgments³ (Tassy, Oullier, Mancini, & Wicker, 2013), and lower approval of aggressive retaliation after being provoked in case of prescriptive, rather than descriptive judgments (Goldstein, Tisak, & Boxer, 2002).

Therefore, we believe that the unexpected results of Huebner and Hauser (2011) can be explained by the type of judgment the participants were asked to make. Specifically, we hypothesize that people would approve of sacrificing themselves more when the judgment was normative, thus exhibiting an other-serving bias, while being less in favor of sacrificing themselves when the judgment was prescriptive, thus exhibiting a self-serving bias. The following studies aim not only to reconcile the currently conflicting empirical results and philosophical intuitions about moral judgments of self-sacrifice, but also to shed more light on how different question prompts may influence moral judgments in general.

Study 1a

Methods⁴

Participants

One hundred and ninety eight German-speaking participants (72% female, median age 23) from the University of Würzburg subject pool took part in the study. The study was a part of several

² Our usage of the term *prescriptive* differs from its usage by some other authors (cf. Janoff-Bulman, Sheikh, & Hepp, 2009). We use the term *prescriptive* to denote what people *should* (not) do. We believe that such judgment is distinct from a judgment of what people are *obligated* (not) to do, which is referred by Janoff-Bulman et al. (2009) as prescriptive judgment. *Normative* judgment is in our conceptualization an answer to the question what behavior is *right* or *wrong*.

³ *Descriptive* judgment denotes what people believe that they or someone else *would* actually do.

⁴ Methods, analyses and hypotheses of the first two studies were pre-registered on Open Science Framework (<https://osf.io/fvd9t/>). The third study was pre-registered before any analysis was conducted but not before the data collection. Analysis scripts, and data can be found at the same place.

unrelated studies administered during a single session. Participants were paid €5 for participation in the whole session, which lasted about 40 minutes.

Materials

We adapted four trolley-like dilemmas from Moore, Clark, and Kane (2008). The dilemmas presented situations where participants had to decide whether to sacrifice one person to save more people. Each dilemma had 2×2 versions differing in whether participants sacrificed themselves or another person (Target of sacrifice: Self vs. Other) and whether they judged if the described action was wrong or if they should perform it given the circumstances (Type of judgment: Wrong – normative vs. Should – prescriptive). For example, one of the dilemmas read as follows (changes in different versions are in parentheses):

You are a crewmember on a marine-research submarine traveling underneath a large iceberg. An onboard explosion has damaged the boat, killed and injured several crewmembers. Additionally, it has caused a collapse of the only access corridor between upper and lower parts of the boat. The upper section, where most of the crew is located, does not have enough remaining oxygen for all to survive until the submarine reaches the surface. Only you are located in the lower section, where there is enough oxygen.

There is an emergency access hatch between the upper and lower sections of the ship. If released by an emergency switch, to which only you have access, it will fall to the deck and allow oxygen to reach the upper section of the submarine.

However, the hatch will crush you, since you are trapped and lying beneath it.
[However, the hatch will crush the crewmember in the lower section, since he is trapped and lying beneath it.]

All of the crew in the upper section of the submarine are almost out of air though, and all will die if you do not do this. On the other hand, if you do nothing, you will survive and will be rescued on the other side of the iceberg.

Should you release the hatch? (prescriptive)
[Is it wrong to release the hatch? (normative)]

Procedure

The experiment was administered in a laboratory setting. Participants did the study on a computer in groups of up to 8. They were separated by barriers and could not interact. Participants first read short instructions that explained the task and then made judgments in the four dilemmas which were presented separately and in a fixed order. Each dilemma had four versions (Target of sacrifice: Self vs. Other × Type of judgment: Wrong vs. Should) and each participant was presented each possible combination exactly once (in a random order). Participants provided answers by clicking on a Yes or No button. The entire experiment was conducted in German.

Results

Data analysis was done using multilevel logistic regression (Gelman & Hill, 2006) with the data for the wrong question reverse coded. Dilemmas and participants were treated as random factors. Effects of the type of judgment, the target of sacrifice, and their interaction were included as predictors. We used a model with random intercepts for both dilemmas and participants, and random slopes for dilemmas.

Results showed that participants were less likely to approve the sacrifice of someone else than themselves, $z = -2.90$, $p = .004$, $OR = 0.30$, 95% CI = [0.13, 0.68], and that they were less likely to say that they should sacrifice the person than that it is not wrong to sacrifice her, $z = -5.43$, $p < .001$, $OR = 0.25$, 95% CI = [0.15, 0.41]. Most importantly, the interaction between the two factors was also significant, $z = 3.00$, $p = .003$, ratio of OR = 3.53, 95% CI = [1.55, 8.05], showing that the self-other asymmetry was more pronounced in the case of normative judgment of wrongness than in the case of prescriptive judgment of what should be done. That can be seen when the analysis was done separately for the two types of judgments. Whereas for the “wrong” question, the sacrificed person effect was significant, $z = -2.92$, $p = .004$, $OR = 0.32$, 95% CI = [0.15, 0.69], it was not significant for the “should” question, $z = 0.26$, $p = .80$, $OR = 1.06$, 95% CI = [0.69, 1.63].

When the analysis was conducted with separate logistic regressions for each dilemma, the interaction between the target of sacrifice and the type of judgment was significant for the last two dilemmas, but not significant for the first two dilemmas.⁵ Separate results for the four dilemmas can be found in Figure 1.

⁵ Results for this analysis can be found on <https://osf.io/rnxp2/>.

Figure 1. Approval of sacrifice by dilemma in Study 1a

The figure displays proportions of participants indicating their approval of sacrificing a person to save more people (i.e., the proportion of people answering “yes” to the “should” question and “no” to the “wrong” question). The order by rows corresponds to the order of presentation of the dilemmas. The error bars represent 95% confidence intervals.

Discussion

The first study showed that there is a difference between normative judgments of wrongness for self- and other-sacrifice. People were more likely to say that it is wrong to sacrifice someone else than themselves. On the other hand, no difference was found in prescriptive judgments of what should be done. The results support our hypothesis that prescriptive and normative moral judgments do not necessarily correspond to each other. However, unlike Huebner and Hauser (2011), we did not observe a self-serving bias in the prescriptive condition.

To replicate the results we conducted a follow-up study with a different sample of participants, which also addressed a possible issue with the order of presentation of the dilemmas.

Specifically, we found a significant interaction between the target of sacrifice and the type of judgment only for the last two dilemmas in Study 1a, so we reversed the order of dilemmas to

examine whether the results could be influenced by their order of presentation (cf. Di Nucci, 2013).

Study 1b

Methods

Participants

Three hundred fifty seven mTurk workers participated in the study. Participants were paid 30¢ for participation, which lasted about 4 minutes.

Materials and procedure

We used the same materials and design as in Study 1a. The only difference being that the dilemmas were presented in English. The experiment was administered online using a custom written web application. Since the effect seemed to be stronger for the last two dilemmas in Study 1a, we reversed the order of the dilemmas. Thus, the last two dilemmas were now the first two dilemmas in this study.

Results

Based on a pre-registered exclusion criterion, 82 participants were removed from the analysis because they answered at least one question in less than 15 seconds. Two further participants were excluded because they did not answer all questions, leaving us with 273 participants.

The same analysis procedure as in Study 1a was used. The results showed again that people were less likely to approve sacrifice of someone else than themselves, $z = -2.73$, $p = .006$, $OR = 0.49$, 95% CI = [0.30, 0.82], and that they were less likely to say that they should sacrifice the person than that it is not wrong to sacrifice her, $z = -3.56$, $p < .001$, $OR = 0.39$, 95% CI = [0.23, 0.65]. Furthermore, the interaction between the two factors was again significant, $z = 2.06$, $p = .04$, ratio of OR = 2.07, 95% CI = [1.04, 4.14], providing additional evidence that the self-other asymmetry is more pronounced in the case of normative judgment of wrongness than in the case of prescriptive judgment of what should be done.

Separate analyses for the two types of judgment again yielded the same results. While there was no effect of the target of the sacrifice when people were asked what should be done, $z = 0.15$, $p = .88$, $OR = 1.03$, 95% CI = [0.70, 1.50], people were less likely to say that it is wrong to sacrifice themselves than someone else, $z = -3.03$, $p = .002$, $OR = 0.51$, 95% CI = [0.33, 0.79].

When the four scenarios were analyzed separately, a significant interaction was found for the first two scenarios (the same two as in the first study).⁶ Separate results for the four dilemmas can be found in Figure 2.

⁶ Results for this analysis can be found on <https://osf.io/v9rgk/>.

Figure 2. Approval of sacrifice by dilemma in Study 1b

The figure displays proportions of participants indicating their willingness to sacrifice a person to save more people (i.e., the proportion of people answering “yes” to the “should” question and “no” to the “wrong” question). The order by rows correspond to the order of presentation of the dilemmas. The error bars represent 95% confidence intervals.

Discussion

Study 1b replicated the results of Study 1a with a different sample of participants and language of materials. We again found the predicted self-other asymmetry. Specifically, while there was no difference in judgment of whether one should sacrifice oneself or someone else, participants were more likely to say that it is wrong to sacrifice someone else than themselves. There was no evidence of an order effect which is in line with recent findings showing that order effects in moral judgment may be relatively uncommon and limited to a narrow set of circumstances (Wiegmann, Okan, & Nagel, 2012). However, the results suggest that the strength of the other-serving bias may depend on used scenarios since we found evidence for the effect only in two dilemmas. Finally, we again did not observe a self-serving bias in the prescriptive condition.

However, one could argue that the obtained results are not caused by different types of judgments, but due to the specific wording that was used. Namely, both studies used negative

framing (“is it wrong?”) for the normative judgment, and positive framing (“should you do it?”) for the prescriptive judgment. The difference between positive and negative framing and not the difference between the two types of judgment might have driven the results. In the next study, we therefore explicitly manipulated the framing in order to address this issue. Alongside the original questions of whether the action is wrong and whether it should be done, we used a positively framed normative judgment (“is it right?”) and a negatively framed prescriptive judgment (“should it not be done?”).

Additionally, we explored whether the results of the first two studies replicate even when the dilemmas were written in the third-person perspective. Therefore, we added a perspective factor, in which we varied whether it was the participants themselves who were deciding in the scenario whether to sacrifice themselves or someone else, or if it was another person named John. Actions and their causes tend to be perceived differently based on whether people perform them themselves or whether they only observe them (Malle, 2006). Previous studies have demonstrated such actor-observer effects also in moral judgment. For instance, Nadelhoffer and Feltz (2008) found that people were more likely to judge as morally permissible to act in a utilitarian way in a trolley dilemma when the scenario was written from the third-person perspective, than when it was written from the first-person perspective – possibly because people experience stronger negative emotions considering the utilitarian sacrifice imagining being an agent rather than observer (Miller, Hannikainen, & Cushman, 2014). We therefore expected that the other-serving bias from the first two studies would be more pronounced in dilemmas presented from the first-person perspective.

Study 2

Methods

Participants

One thousand, eight hundred and ninety two participants were given a pen-and-paper questionnaire (which asked participants mainly about university and high school education) before the General Academic Prerequisites (GAP) test. The GAP test is used for university admission at the undergraduate level by a wide range of universities in the Czech Republic. The scenario used in the present study was included in the questionnaire. We excluded 314 participants for failing to respond to the question, leaving us with 1578 participants. In the remaining sample, 54.7% of participants were women and the median age was 19.4 years (*IQR* = 1.49).

Materials and procedure

Only the submarine dilemma scenario from the previous studies was used.⁷ There were 16 versions of the dilemma in total, resulting from a combination of four factors. In addition to the four versions that were the same as in Study 1 (i.e., Type of judgment: Normative vs. Prescriptive, and Target of sacrifice: Self vs. Other) we added the factor of framing (Positive vs. Negative) and the perspective factor (“You” vs. “John”). The participants were randomly

⁷ The wording was slightly adjusted so that it made grammatical sense in all the conditions, see <https://osf.io/n7e8c/>.

assigned to a single version of the scenario and provided their responses using a six-point Likert scale ranging from 1 (completely agree) to 6 (completely disagree). Responses for the positive framing scenarios were reversed such that higher values always corresponded to higher approval of the sacrifice (i.e., similarly as in the first two studies). All the predictor variables were mean centered.

Because we also had available the results of the GAP test for the participants, we used them for exploratory analyses of the effect of cognitive ability on evaluation of the moral dilemmas. The test is similar in structure to the Scholastic aptitude test (SAT; for additional details, see Bahník & Vranka, 2017) used for university admissions in the US, which was shown to correlate highly with general intelligence (Frey & Detterman, 2004).

Results⁸

Full model⁹

All four factors as well as all their interactions were inserted into a regression model. The full results are presented in Table 1 (see also Figure 3). They show that people approved of sacrifice more in the positive frame ($M = 4.16$, $SD = 1.18$) than in the negative frame ($M = 3.68$, $SD = 1.31$), and that participants were more willing to sacrifice someone else ($M = 4.01$, $SD = 1.21$) than themselves ($M = 3.84$, $SD = 1.31$), suggesting a general self-serving bias. Most importantly, the model revealed a significant three-way interaction between the frame, target of sacrifice, and type of judgment so we next decomposed the interaction by frames. Since there was no significant effect of the perspective factor, the results are pooled across both perspectives.

⁸ A direct replication of Study 1 showed an interaction of judgment type and target of sacrifice in the same direction as in Study 1, which was however not significant (see <https://osf.io/2kpnh/> for full results).

⁹ The manner in which we analyze the data in Study 2 is slightly different from our pre-registration. We separated the four questions by two factors (framing and judgment type) instead of treating them as a single factor with four levels as preregistered. We considered this analysis to be more in line with the way the analysis was conducted in Study 1, thus facilitating comparisons between the studies.

Figure 3. Approval of sacrifice by condition in Study 2

The figure displays the average rating of the approval to sacrifice one person to save more people (i.e., reverse scored agreement with negatively framed questions). The error bars represent 95% confidence intervals. Individual data with applied jitter are shown as transparent points.

	Full model (N = 1578)		Positive frame (N = 809)		Negative frame (N = 769)	
	<i>b</i>	95% CI	<i>b</i>	95% CI	<i>b</i>	95% CI
(Intercept)	3.93**	[3.87, 3.99]	4.16**	[4.08, 4.24]	3.68**	[3.60, 3.77]
Framing	0.48**	[0.35, 0.60]				
Judgment Type	0.08	[-0.04, 0.21]	0.09	[-0.07, 0.25]	0.07	[-0.11, 0.26]
Perspective	-0.04	[-0.17, 0.08]				
Sacrifice Target	-0.18**	[-0.30, -0.06]	-0.24**	[-0.40, -0.08]	-0.11	[-0.30, 0.07]
Framing * Judgment Type	0.01	[-0.23, 0.26]				
Framing * Perspective	-0.12	[-0.37, 0.12]				
Judgment Type * Perspective	-0.17	[-0.41, 0.08]				

<i>Framing * Sacrifice Target</i>	-0.13	[-0.37, 0.12]				
<i>Judgment Type * Sacrifice Target</i>	0.11	[-0.14, 0.35]	-0.19	[-0.51, 0.14]	0.43*	[0.06, 0.80]
<i>Perspective * Sacrifice Target</i>	0.09	[-0.15, 0.34]				
<i>Framing * Judgment Type * Perspective</i>	0.01	[-0.48, 0.50]				
<i>Framing * Judgment Type * Sacrifice Target</i>	-0.62*	[-1.11, -0.13]				
<i>Framing * Perspective * Sacrifice Target</i>	-0.11	[-0.60, 0.38]				
<i>Judgment Type * Perspective * Sacrifice Target</i>	0.34	[-0.15, 0.83]				
<i>Framing * Judgment Type * Perspective * Sacrifice Target</i>	-0.40	[-1.38, 0.58]				

Table 1. Regression coefficients and 95% confidence intervals for the full model and two models separated by the frame factor for Study 2. Note: * indicates $p < .05$; ** indicates $p < .01$.

Positive frame

We first took a closer look at the condition where the questions were positively framed. Similar to Study 1, we focused on how the effect of type of judgment (Prescriptive vs. Normative) might differ based on the target of judgment (Self vs. Other). The results indicated that, in the positive frame, participants approved more of sacrificing the other person ($M = 4.27$, $SD = 1.11$) than themselves ($M = 4.04$, $SD = 1.24$), suggesting a self serving bias, irrespective of whether the judgment was described as prescriptive or normative (see Table 1). Cognitive ability, as measured by the GAP test, did not moderate any of the results in the positive frame.

Negative frame

We then took a closer look at the condition where the questions were negatively framed. Similar to above, we focused on how the effect of type of judgment (Prescriptive vs. Normative) might differ based on the target of judgment (Self vs. Other). The results indicated an interaction between the type of judgment and the target of sacrifice (see Table 1). Separate analyses for the two types of judgments show that, when the judgment was prescriptive, participants agreed more that they should not sacrifice themselves ($M = 3.48$, $SD = 1.37$) than the other person ($M = 3.81$, $SD = 1.25$), $b = -0.34$, $p = .014$, 95% CI [-0.60, -0.07], suggesting a self-serving bias. When the judgment was normative, on the other hand, participants agreed more that it is wrong to sacrifice someone else ($M = 3.67$, $SD = 1.26$) than themselves ($M = 3.80$, $SD = 1.33$). However, this difference, although in the same direction as in Study 1, was not significant, $b = 0.10$, $p = .45$, 95% CI [-0.16, 0.35]. Even though the interaction between the target of sacrifice and perspective was not significant, we next analyzed the effect of target of sacrifice only for the first person perspective for the normative judgment in the negative frame. Because this condition corresponds directly to the materials used in Study 1, this amounted to a direct replication of the “wrong” condition. The results show that participants agreed more that it is wrong to sacrifice someone else ($M = 3.54$, $SD = 1.26$) than themselves ($M = 3.84$, $SD = 1.32$), but the difference was not significant, $b = 0.30$, $p = .10$, 95% CI [-0.06, 0.65] (see Figure 3). Additionally, unlike in the positive frame, when the questions were negatively framed,

participants with a higher cognitive ability displayed generally higher approval of sacrifice, $r = .14$, $p < .001$, 95% CI [0.07, 0.21].

Discussion

As in previous studies, we again found evidence of self-other asymmetry, although not in accord with all our predictions. Overall, participants showed a general self-serving bias. More interestingly though, the addition of the framing factor proved to have a significant impact on the results. Separate analyses of the two frames showed that when the questions were positively framed, participants displayed a self-serving bias in their judgments. That is, they were more likely to approve of sacrificing someone else than themselves, independent of the type of judgment. When the questions were negatively framed, participants were more likely to exhibit a self-serving bias when the judgment was prescriptive. On the other hand, when the judgment was normative, there was a similar tendency for an other-serving bias as in Study 1. However, the effect was not significant – possibly due to lower statistical power or the use of a different response scale. These results support and extend the findings from Study 1.

The fact that we did not observe a self-serving bias in previous experiments might have been caused by using only a positively framed prescriptive condition in which the bias is less pronounced even in the current study (see Figure 3). While the two types of judgments produced different effects, it must be highlighted that the differences seemed to be qualified by framing. Framing and judgment type seem to be interdependent and future investigations of moral judgments in general should take this into account.

Surprisingly, there was no significant effect of perspective on moral evaluations – whether it was the participants themselves or someone else (i.e., “John”) who were making the decision in the scenario did not seem to make a difference. Our specific prediction was that the other-serving bias would be more pronounced in the first person condition. As can be seen in Figure 3, the tendency for an other-serving bias was observed only in the first person perspective. However, the interaction with the perspective factor was not significant, again possibly because of low statistical power.

Finally, we conducted exploratory analyses of possible moderation by cognitive ability, as measured by the GAP test. While cognitive ability did not seem to have an impact when the judgments were positively framed, higher cognitive ability was associated with higher willingness to sacrifice in the negative frame, irrespective of judgment type and target of sacrifice. These results are partially in line with Greene’s (2012) argument that since utilitarian decisions are based on reason, while deontologic are based on emotion, people with higher cognitive ability should be more likely to make utilitarian judgments, i.e., in our case, to be more inclined towards sacrifice. It is however not clear why the effect is missing in the positive frame. One possibility is that framing the question negatively focuses participants on reasons against the sacrifice – as suggested by the lower overall approval in the negative frame. Higher cognitive ability is then needed in order to overcome this effect (Moore et al., 2008) – however, in the positive frame, there is no such need as participants are more likely to be focused on reasons in favor of the sacrifice from the beginning.

General Discussion

We are all familiar with the trope of the hero sacrificing himself or herself for the good of the many. One intuitively expects that, in certain situations, people would agree to sacrifice themselves rather than others – thus exhibiting an other-serving bias. Indeed, there have been suggestions that this is not only morally permissible, but also praiseworthy (Splawn, 2001). Similarly, Chisholm (1963) suggested that accepting physical harm to oneself in order to benefit others has sometimes been considered as supererogatory, that is morally praiseworthy although not required. However, in previous studies that looked at moral judgments, researchers were mostly focused on, and have quite often been finding, examples of self-serving biases – even in situations when the opposite other-serving bias would have been intuitively expected (Huebner & Hauser, 2011).

In order to explore some of these discrepancies between intuitions and empirical findings, in this paper we took a closer look at one factor that might be particularly important for studying self-other asymmetries in moral judgments – namely, whether the judgment in question was prescriptive (i.e., whether something should or should not be done) or normative (i.e., whether something is right or wrong). In three experiments, we found that people were more likely to approve of sacrificing themselves than others only when the judgment was framed negatively and in a normative way; that is, when participants were asked whether it is wrong to sacrifice themselves or someone else. Our results thus show an example of an other-serving bias in moral judgments and provide empirical support for intuitions described by Slote (1984) and Stocker (1976). However, it is worth highlighting that the bias in this direction seems to appear only in a very narrow set of conditions (i.e., only when judging whether the sacrifice is wrong) and is probably less common than would be expected. This is in line with recent findings of low willingness to undergo harm for benefit of others in laboratory settings (Volz, Welborn, Gobel, Gazzaniga, & Grafton, 2017). Additionally, in accordance with the results of previous studies (e.g., Ditto, Pizarro, & Tannenbaum, 2009; Huebner & Hauser, 2011; Valdesolo & DeSteno, 2007; 2008), in Study 2 we also observed the much more common self-serving bias, especially when participants judged whether the sacrifice is right and whether it should not be done.

The finding of an other-serving bias only in normative judgment is in accord with the claim by Janoff-Bulman and colleagues (2009) that people employ at least partially different processes when making prescriptive and normative judgments. Namely, people could be focusing more on rules and prohibitions when deciding whether an action is wrong. Their normative moral judgment would therefore be more deontological and less concerned with evaluation of consequences. In the present case, people would focus mainly on harm done to others, because they see normative moral rules as applying mostly to harm done to someone else (Gray & Wegner, 2011). Therefore, they do not view self-sacrifice as a rule violation and judge it as not wrong. On the other hand, when thinking about what should and should not be done, these rules could be less salient thus making consideration of other factors such as consequences and self-interest more likely to affect the judgment.

The difference between descriptive and normative judgment found in the study by Tassy and colleagues (2013) is generally in accord with the proposed explanation. The study found that

people are more willing to approve of sacrificing a person to save more people when making a descriptive judgment (“would you?”) than normative judgment (“is it acceptable?”). Additionally, Tassy and colleagues found that affective proximity to the sacrificed person had larger effect on descriptive judgment. It is possible that people are less influenced by affective proximity in normative judgment because the moral rule not to kill is the same for strangers as well as for close relatives. However, the consequences of killing one’s relative or a stranger are different. And, because consequences are possibly more salient when judging what one would do (similarly as when judging what should be done), the difference in consequences influences the descriptive judgment more. Similarly, people making prescriptive judgments in our studies might have taken into consideration not only rules forbidding killing of others, but their own self-interest as well.

However, these explanations do not account for the unexpected lack of other-serving bias in the positively framed normative judgment, that is when participants judged whether the sacrifice is right. Although we originally expected the effect to be present for both judgments of wrong and right, framing effects on moral judgments have been observed before (e.g., Petrinovich & O’Neill, 1996; Pastötter et al., 2013). One possible explanation why we did not observe an other-serving bias in positively framed normative judgment could be the lack of positive deontological rules similar to those prohibiting harming others – as mentioned by Chisholm (1963), suffering harm for the benefit of others is considered praiseworthy, but not obligatory. In a case of positively framed normative judgment, the deontological intuition, responsible for the other-serving bias observed in the negative frame, is therefore absent, allowing other considerations such as self-interest to affect the judgment.

The observed effects of judgment type and framing underscore how question wording may influence moral judgment (O’Hara, Sinnott-Armstrong, & Sinnott-Armstrong, 2010; Pastötter, Gleixner, Neuhauser, & Bäuml, 2013; Tassy et al., 2013). It is possible that similar wording effects might limit generalizability of other previously found effects on moral judgment and researchers should pay attention to wording effects in future studies to overcome these possible limitations.

Interestingly, we did not find an effect of perspective on judgments (i.e., whether the actor in the scenario was the participant or someone else). Nadelhoffer and Feltz (2008) demonstrated a type of “actor-observer” bias where people hold other people to different moral standards than themselves. We were also interested in this effect because it can be considered another example of a self-other asymmetry. Nadelhoffer and Feltz argue that the explanation of this effect relies on different affective responses people have in scenarios where they are the actor or observer. It could be that our scenarios failed to evoke these affective reactions. However, our failure to obtain this effect is not an isolated one. Recently, Boudesseul and Lantian (2017) and Andow (2017) failed to replicate this effect in a high powered direct and conceptual replication attempts. While there might be some subtle differences between our study and the original, the findings that we obtained here in combination with the recently failed replication attempts invite caution in interpreting the effect of perspective on moral judgments.

A possible limitation of our experiments is that we used only a few scenarios in the first study and only one in the second. Although we use a random effect for stimuli in analysis of Study 1, we were unable to reliably examine potential effects of scenarios on the judgment. When only a fixed set of stimuli is used in a study, the generalizability of its findings is always unknown (Bahník & Vranka, 2017) and stimuli sampling, when possible, is therefore recommended.

The goal of the present study was to explore how philosophical intuitions about moral judgments of self-sacrifice can be reconciled with conflicting empirical findings. In this vein, we looked at whether self-other asymmetries in judgments of trolley-like dilemmas are affected by whether the judgment is normative or prescriptive. The results of three experiments where we manipulated the type of judgment, its framing, and the perspective from which the dilemmas were presented, showed that an other-serving bias emerges only for negatively framed normative judgment while a self-serving bias was observed under all remaining conditions. Taken together, the findings appear to be quite robust as they were obtained across three different samples from different cultural milieus (although all belonging to the Western culture). However, future work could focus on identifying other instances where self-other asymmetries might be observed, how different judgment types might impact this tendency, and whether these biases can be replicated in real-world settings.

Acknowledgements

We would like to thank Ira Theresa Maschmann for translating materials for Study 1a and administering it, Anita Körner for helpful discussion, Jana Dlouhá for creating a web application for Study 1b, and Dale Barr for statistical advice. We would like to especially thank Scio company for the data from the National Comparative Exams GAP test used in Study 2. The research was partly supported by Deutsche Forschungsgemeinschaft (DFG-RTG 1253/2). The work of Štěpán Bahník and Marek Vranka was supported by Internal Grant Agency of Faculty of Business Administration, University of Economics, Prague (IP300040).

References

- Alloy, L. B., & Ahrens, A. H. (1987). Depression and pessimism for the future: biased use of statistically relevant information in predictions for self versus others. *Journal of personality and social psychology*, 52(2), 366-378.
- Andow, J. (2017). Are Intuitions About Moral Relevance Susceptible to Framing Effects? *Review of Philosophy and Psychology*, 1-27.
- Bahník, Š., & Vranka, M. A. (2017). If it's difficult to pronounce, it might not be risky: The effect of fluency on judgment of risk does not generalize to new stimuli. *Psychological science*, 28(4), 427-436.
- Barbosa, S., & Jiménez-Leal, W. (2017). It's not right but it's permitted: Wording effects in moral judgement. *Judgment and Decision Making*, 12(3), 308-313.

- Chisholm, R. M. (1963). Supererogation and offence: a conceptual scheme for ethics. In R. M. Chisholm, Brentano and Meinong Studies (pp. 98-113), Atlantic Highlands, NJ: Humanities Press.
- Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences*, 111(48), 17320-17325.
- Di Nucci, E. (2013). Self-Sacrifice and the Trolley Problem. *Philosophical Psychology*, 26(5), 662–672.
- Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. *Psychology of learning and motivation*, 50, 307–338.
- Epley, N., & Dunning, D. (2000). Feeling "holier than thou": are self-serving assessments produced by errors in self-or social prediction? *Journal of personality and social psychology*, 79(6), 861-875.
- Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, 145(6), 772-787.
- Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, 5, 5–15.
- Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological science*, 15(6), 373-378.
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.
- Gray, K., & Wegner, D. M. (2011). Morality takes two: Dyadic morality and mind perception. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 109–127). Washington, DC: APA.
- Greene, J. D. (2007). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development* (pp. 35–80). Cambridge, MA: MIT Press.
- Greene, J. D. (2012). *The moral brain and how to use it*. New York: Penguin.
- Hoorens, V. (1995). Self-favoring biases, self-presentation, and the self-other asymmetry in social comparison. *Journal of Personality*, 63(4), 793-817.
- Huebner, B. & Hauser, M. D. (2011). Moral judgments about altruistic self-sacrifice: When philosophical and folk intuitions clash. *Philosophical Psychology*, 24(1), 73–94.
- Janoff-Bulman, R., Sheikh, S. & Hepp, S. (2009). Proscriptive versus prescriptive morality: two faces of moral regulation. *Journal of Personality and Social Psychology*, 96(3), 521–539.

- Lantian, A., Boudesseul, J., & Cova, F. (2017, July 26). XPhi Replicability Project - Replication of Nadelhoffer & Feltz (2008, Neuroethics). Retrieved from osf.io/jb8yp
- Lester, D. (2014). Altruistic suicide in the trolley problem. *Comprehensive Psychology*, 3, 6.
- Malle, B. F. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychological bulletin*, 132(6), 895-919.
- Miller, R. M., Hannikainen, I. A. & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion*, 14(3), 573–587.
- 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(6), 549–557.
- O'Hara, R. E., Sinnott-Armstrong, W. & Sinnott-Armstrong, N. A. (2010). Wording effects in moral judgments. *Judgment & Decision Making*, 5(7), 547–554.
- Pastötter, B., Gleixner, S., Neuhauser, T., & Bäuml, K. H. T. (2013). To push or not to push? Affective influences on moral judgment depend on decision frame. *Cognition*, 126(3), 373–377.
- Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, 36(1), 163-177.
- Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, 17(3), 145-171.
- Portmore, D. W. (2001). Can an act-consequentialist theory be agent relative? *American Philosophical Quarterly* 38(4), 363–377.
- Rohsenow, D. J. (1983). Drinking habits and expectancies about alcohol's effects for self versus others. *Journal of Consulting and Clinical Psychology*, 51(5), 752-756.
- Rom, S. C., & Conway, P. (2018). The strategic moral self: Self-presentation shapes moral dilemma judgments. *Journal of Experimental Social Psychology*, 74, 24-37.
- Splawn, C. (2001). The Self-other Asymmetry and Act-utilitarianism. *Utilitas*, 13(03), 323–333.
- Slote, M. (1984). Morality and Self-Other Asymmetry. *The Journal of Philosophy*, 81(4), 179–192.
- Stocker, M. (1976). Agent and other: Against ethical universalism. *Australasian Journal of Philosophy*, 54(3), 206–220.
- Tassy, S., Oullier, O., Mancini, J. & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology*, 4, 250.

Uhlmann, E. L., Zhu, L. L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition*, 126(2), 326-334.

Valdesolo, P., & DeSteno, D. A. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, 18(8), 689–690.

Valdesolo, P. & DeSteno, D. (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology*, 44(5), 1334–1338.

Volz, L. J., Welborn, B. L., Gobel, M. S., Gazzaniga, M. S., & Grafton, S. T. (2017). Harm to self outweighs benefit to others in moral decision making. *Proceedings of the National Academy of Sciences*, 2017066693.

Wiegmann, A., Okan, Y., & Nagel, J. (2012). Order effects in moral judgment. *Philosophical Psychology*, 25(6), 813-836.