The power of moral words: Loaded language generates framing effects in the extreme dictator game

Valerio Capraro*   Andrea Vanzo†

Abstract

Understanding whether preferences are sensitive to the frame has been a major topic of debate in the last decades. For example, several works have explored whether the dictator game in the give frame gives rise to a different rate of pro-sociality than the same game in the take frame, leading to mixed results. Here we contribute to this debate with two experiments. In Study 1 (N=567) we implement an extreme dictator game in which the dictator either gets $0.50 and the recipient gets nothing, or the opposite (i.e., the recipient gets $0.50 and the dictator gets nothing). We experimentally manipulate the words describing the available actions using six terms, from very negative (e.g., stealing) to very positive (e.g., donating) connotations. We find that the rate of pro-sociality is affected by the words used to describe the available actions. In Study 2 (N=221) we ask brand new participants to rate each of the words used in Study 1 from "extremely wrong" to "extremely right". We find that these moral judgments can explain the framing effect in Study 1. In sum, our studies provide evidence that framing effects in an extreme Dictator game can be generated using morally loaded language.

Keywords: framing effect, moral preferences, dictator game, moral judgment

1 Introduction

Understanding whether preferences are sensitive to non-economic cues, such as the name that the experimenter gives to the game or the words used to describe the available actions, has been a topic of debate in the last two decades.

Earlier works found that people cooperate more in the prisoner's dilemma when it is called Community Game than when it is called Wall Street Game (Kay & Ross, 2003; Liberman et al., 2004). Yet, it was soon noticed that the presence of social framing effects in the prisoner's dilemma does not imply that preferences are sensitive to context. It might be that social framing works as a coordination device, by affecting participants' beliefs about the behavior of the other people involved in the interaction (Fehr & Schmidt, 2006). Consistent with this view, Ellingsen et al. (2012) found that social framing effects in the prisoner's dilemma vanish when the game is played sequentially, suggesting that social cues primarily works by changing participants' beliefs.

To shed light on this issue, in more recent years scholars have started investigating whether social framing can affect people's choices in social decision problems, as opposed to strategic games. In this context, the decision maker can affect the outcome of one or more other people, but these people have no way to affect the decision maker's outcome. Since in these situations beliefs play no role, any framing effect would be uniquely driven by a chance in preferences.

A stream of research has focused on whether the words used to describe the available actions can affect decision makers' choices. The empirical evidence has shown that indeed they can, at least in some contexts. Eriksson et al. (2017) have found that rejection rates in the Ultimatum game depend on whether the rejection option is called "reduction of the other player's payoff" or "rejection of the other player's offer". In the former case, responders are less likely to reject the proposer's offer, and this effect appears to be driven by moral concerns. Similarly, Capraro and Rand (2018) and Tappin and Capraro (2018) have found that behavior in the Trade-Off game, i.e., a decision problem in which decision-makers have to decide between an equitable and an efficient allocation of resources, highly depends on which choice is presented as being the "right thing to do".

Particularly interesting is the situation regarding the Dictator game (DG). The DG, whereby the *dictator* has to decide how to unilaterally divide a sum of money between her/himself and the *recipient*, is considered to be the standard measure of altruistic behavior and, for this reason, is among the most studied decision problems in social science (Forsythe et al., 1994; Eckel & Grossman, 1996; Engel, 2011; Rand et al., 2016). Studies exploring framing effects in the DG have led to mixed results. Previous works have mainly focused on whether the DG in the *Take* frame gives rise to higher pro-sociality than the same game in the *Give* frame. Swope et al. (2008) have found that participants tend to be more pro-social in the DG in the *Take* frame than in the DG in the *Give* frame. In the *Take* frame, the endowment

*Middlesex University London. Email: V.Capraro@mdx.ac.uk.
†Heriot-Watt University Edinburgh. Email: a.vanzo@hw.ac.uk

was initially given to the recipient, and dictators could take any amount. Krupka and Weber (2013) have replicated this framing effect in a context in which the initial endowment is split equally between the dictator and the recipient. Additionally, Krupka and Weber (2013) found that this framing effect is driven by a change in the perception of what is the "socially appropriate thing to do". However, the framing effect when passing from the *Take* frame to the *Give* frame has not been replicated by several other works (Dreber et al., 2013; Grossman & Eckel, 2015; Halvorsen, 2015; Hauge et al., 2016; Goerg et al., 2017), casting doubts on the very existence of framing effects in the DG.[1]

Here we focus on the existence of framing effects in the DG with two experiments. In Study 1, we implement an extreme DG, where the dictator can either allocate $0.50 to himself and $0 to the recipient, or conversely, $0.50 to the recipient and $0 to himself. We then manipulate the words used to describe the two available actions through six different frames. We find that this word manipulation significantly affects participants' decisions. In Study 2 we ask brand new participants to rate all the words used to describe the available actions in Study 1, from "extremely wrong" to "extremely right". We show that the rate of pro-sociality in a given frame of Study 1 can be predicted by the difference between the rating of the word associated, in the same frame, with the pro-social action and the rating of the word associated, in the same frame, with the pro-self action.

2   Study 1

We consider a variant of the standard DG. In the standard DG, the dictator has to allocate an amount of money (e.g., $0.50) between her/himself and another participant (typically anonymous). The recipient has no choice and only gets what the dictator decides to give. Conversely, in our *extreme* version, the dictator can opt only for the two extreme options: either s/he gets the whole $0.50 (and the recipient gets nothing) or the recipient gets the $0.50 and the dictator gets nothing. The reason why we choose this extreme variant, rather than the standard one, is to be able to write the instructions for the decision problem in several different treatments by changing just one word. This would have been hard, if not impossible, with the classical "continuous" variant, as it will be clear later. In this extreme DG, we hypothesize that the words used to describe the available actions will affect participants' choices beyond the economic consequences that

these actions bring about. Under the same economic conditions, naming the self-regarding action through an extremely "negative" expression as *Steal from the other participant*, rather than using the more "neutral" expression *Take from the other participant*, should reduce the frequency of the selfish choice.

2.1   Experimental Design

Participants are randomly assigned to one of twelve conditions. In the *Steal vs Don't steal* condition, they are told that there are $0.50 available and they have to choose between two possible actions: *Steal from the other participant*, so that they would get the $0.50 and the other participant would get nothing; or *Don't steal from the other participant*, so that the other participant would get the $0.50, while they would get nothing. It is made explicit that the other participant has no choice and will be really paid according to the decision made. Participants are also asked two comprehension questions, to make sure they understand the decision problem. One question asks which choice would maximize their payoff; the other one asks which choice would maximize the other participant's payoff. Participants failing either or both comprehension questions are automatically excluded from the survey. Those who pass the comprehension questions are asked to make the real choice. The *Don't steal vs Steal* condition is similar: the only difference is that we switched the order of presentation of the options, in order to account for order effects. The conditions *Take vs Don't take*, *Don't take vs Take*, *Demand vs Don't demand*, *Don't demand vs Demand*, *Give vs Don't give*, *Don't give vs Give*, *Donate vs Don't donate*, *Don't donate vs Donate*, *Boost vs Don't boost*, *Don't boost vs Boost* are analogous. Hence, the instructions of one frame differ from those of another frame only in one word (see Appendix for full instructions). This justifies our choice of an extreme variant of the Dictator game, rather than the classical "continuous" variant. A moment of reflection shows that it is very hard, if not impossible, to write, for example, the *Take* frame and the *Give* frame in a "continuous" form using instructions differing only in one word.[2]

A standard demographic questionnaire concludes the survey. At the end of the questionnaire, participants are given the completion code with which they can submit the survey to Amazon Mechanical Turk (AMT) and claim their

---

[1]Related to this literature, some studies have shown that extending the choice set of the standard DG to include a taking option has the effect of decreasing giving (List, 2007; Bardsley, 2008; Cappelen et al., 2013; Zhang & Ortmann, 2013). Eichenberger and Oberholzer-Gee (1998) have found that the amount *given* by students who perform well in a test to students who perform bad in the same test is greater than the amount *not taken* by students who perform bad in a test from students who perform well. Chang, Chen and Krupka (2017) have found that the DG in the tax frame gives rise to more pro-sociality than the standard DG.

[2]The choice of the words "steal", "take", "demand", "donate", "give", and "boost" was motivated by Sentiment Analysis (Pang & Lee, 2004; Pang et al., 2002; Liu et al., 2007; McGlohon et al., 2010; Vanzo et al., 2014). In particular, we based our choice on a lexical resource known as SentiWordNet (Baccianella et al., 2010; Esuli & Sebastiani, 2006), which divides the English vocabulary in synsets (i.e., sets of synonyms). We selected six synsets containing words that are particularly suitable to describe actions in an extreme dictator game: ***steal#1*** = "take without the owner's consent"; ***take#8*** = "take into one's possession"; ***demand#1*** = "request urgently and forcefully"; ***give#3*** = "transfer possession of something concrete or abstract to somebody"; ***donate#1*** = "give to a charity or good cause"; and ***boost#2*** = "be beneficial to".

payment. Payoffs are computed and paid on top of their participation fee ($0.50). The other participant was selected at random from the same sample. Therefore, participants received a payment both from their choice as decision makers, and from being in the role of the "other participant" for a different participant. When making their choice, participants were not informed that they would also be in the role of the "other participant" to avoid having this affect their decision. The Appendix shows verbatim experimental instructions.

2.2 Results

2.2.1 Participants

We recruited US based participants on the online platform Amazon Mechanical Turk (AMT).[3] We collected 727 observations and cleaned this dataset through the following two operations. First, we looked for multiple observations by checking multiple IP addresses and multiple Turk IDs. In case we found multiple observations, we kept only the first one as determined by the starting date and we deleted all the remaining ones. Second, we discarded all participants who failed either or both comprehension questions. After these operations, we remained with a sample of $N = 567$ participants (52.7% females, mean age = 37.6 years). Thus, in total, 22% of the participants have been eliminated from the analysis. This is in line with previously published studies using AMT (Horton et al., 2011).

2.2.2 Pro-Sociality

To analyze this sample, we first build a binary variable, named *Pro-Sociality*, which takes value 1 whenever a participant allocates the $0.50 to the other person, and 0 whenever a participant allocates the $0.50 to him/herself. Average *Pro-Sociality* across the whole sample is 0.141, that is, 14.1% of the participants act pro-socially. *Pro-Sociality* within a frame turns out to not depend on the order in which the options are presented, e.g., average *Pro-Sociality* in the condition *Steal vs Don't steal* is not statistically different from average *Pro-Sociality* in the condition *Don't steal vs Steal*. Similarly for all other frames. Thus, in what follows we collapse across order and we name w the union between condition w vs *Don't w* and *Don't w* vs w.

2.2.3 Framing effect

We now test our hypothesis that the words used to describe the available actions affect *Pro-Sociality*. Figure 1 provides visual evidence that there is significant variability across conditions: the minimum of the variable *Pro-Sociality* is

---

[3]Several works have shown that data gathered using AMT are of no less quality than data collected using the standard laboratory (Arechar et al., 2018; Brañas-Garza et al., 2018; Horton et al., 2011; Paolacci & Chandler, 2014; Paolacci et al., 2010; Rand, 2012).
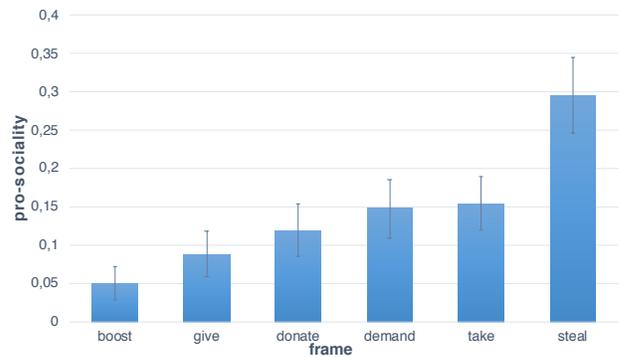

FIGURE 1: *Pro-Sociality* across frames in Study 1. Error bars represent standard errors of the means.

TABLE 1: Pairwise logit regression predicting the effect on *Pro-Sociality* of passing from one frame of Study 1 to the other one. We report coefficients and, in brackets, standard errors. Significance thresholds: *: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$.

	Give	Donate	Demand	Take	Steal
Boost	−0.607	−0.937*	−1.181**	−1.234**	−2.065***
	(0.590)	(0.560)	(0.548)	(0.529)	(0.515)
Give		−0.331	−0.575	−0.628	−1.458***
		(0.490)	(0.477)	(0.455)	(0.438)
Donate			−0.244	−0.297	−1.128***
			(0.440)	(0.416)	(0.397)
Demand				−0.053	−0.884**
				(0.400)	(0.381)
Take					−0.830**
					(0.352)

attained in the *Boost* frame (5.0%), while its maximum is attained in the *Steal* frame (29.5%). Table 1 reports coefficients and standard errors of pairwise logit regressions. In short, it emerges that the *Steal* frame produces an amount of pro-sociality that is statistically significantly higher than that of any other frames. Conversely, the *Boost* frame gives rise to an amount of pro-sociality that is numerically lower than that of any other frames; however, the difference is statistically significant only versus the *Demand*, *Take*, and *Steal* frames, while it is marginally significant versus the *Donate* frame. The other frames numerically lie between the *Boost* frame and the *Steal* frame, but pairwise differences are not statistically significant.

3 Study 2

Study 1 shows that the words used to describe the available actions in an extreme dictator game can affect participants' decisions. How can we explain this result? Since different frames have the same economic structure, this finding is inconsistent with simple economic theories of social preferences, which assume that people's utility function depends only on the economic consequences of the available actions (Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000; Charness & Rabin, 2002; Engelmann & Strobel, 2004). One potential explanation is that participants have moral preferences: they compare the moral goodness of the pro-social action to the moral badness of the pro-self action, and then act pro-socially only when the resulting difference is large enough to counterbalance the cost of the pro-social action. This explanation makes the empirically testable prediction that there is a correlation between pro-sociality in frame w and the difference between the moral judgment associated to the pro-social action in frame w and the pro-self action in frame w. In Study 2 we test this hypothesis.

3.1 Experimental design

Participants are presented, in random order, the instructions of all the six frames of the extreme Dictator game in Study 1. After reading the instructions of the extreme Dictator game, for each frame word $w \in \{Steal, Demand, Take, Give, Boost, Donate\}$, e.g. $w = Steal$, participants are asked the following two questions in random order:

- *From a moral point of view, how would you judge the choice: to steal?*
- *From a moral point of view, how would you judge the choice: not to steal?*

Answers are collected using a 5-point Likert scale with: 1 = "extremely wrong", 2 = "somewhat wrong", 3 = "neutral", 4 = "somewhat right", and 5 = "extremely right". After answering all these twelve (two for each frame) questions, participants enter the demographic questionnaire ending the survey. Note that, in this study, participants do not make any decision.[4]

3.2 Results

3.2.1 Participants

We recruited 250 US based participants on AMT. None of them participated in Study 1. After eliminating multiple IP addresses and multiple TurkIDs, we remained with 221 observations (female = 41.9%, mean age = 34.25).

---

[4]We deliberately decided to collect moral judgments using a different sample to avoid that any correlation be driven by subjects justifying their choice in terms of morality.
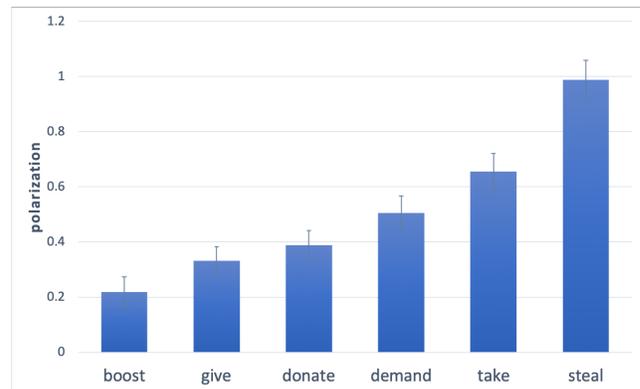

FIGURE 2: *Polarization* across conditions in Study 2. Error bars represent standard errors of the means.

3.2.2 Predicting pro-sociality from self-reported moral judgments

We aim at showing that the rate of pro-sociality in Study 1 in frame w parallels the difference between the moral judgment of the pro-social action in frame w and the moral judgment of the pro-self action in frame w.

To this end, we define the following sets. Let $W^+ = \{Don't\ steal, Don't\ take, Don't\ demand, Boost, Give, Donate\}$ denote the set of wordings corresponding to the pro-social action. Similarly, let us also define the set $W^- = \{Steal, Take, Demand, Don't\ boost, Don't\ give, Don't\ donate\}$, to be the set of wordings corresponding to the pro-self action. For each $w \in W^+$, let $\neg w$ be the corresponding word in W^-, that is, for example, $\neg(Don't\ steal) = Steal$.

For each $w \in W^+$, we define the polarization of w to be $Pol(w) = Judgm(w) - Judgm(\neg w)$, where $Judgm(w)$ is the self-reported moral judgment of w, collected in Study 2 and normalized between -1 and 1, with 0 corresponding to "neutral".

Figure 2 reports all the polarizations. When comparing it to Figure 1, one immediately notices that the pattern of the polarizations is very similar to the pattern of pro-sociality.

4 Discussion

In Study 1 we have shown that the words used to describe the available actions in an extreme Dictator game can significantly alter subjects' behavior. Study 2 shows that this framing effect is consistent with a theory of moral preferences, according to which participants compare the moral goodness of the pro-social action to the moral badness of the pro-self action, and then act pro-socially only when their difference is large enough to counterbalance the cost of the pro-social action. In this view, the words in Study 1 affect behavioral outcomes by affecting the moral judgments associated to the available actions.

A number of papers have explored whether the Dictator Game in the *Take* frame gives rise to greater pro-sociality than the same game in the *Give* frame. Swope et al. (2008) and Krupka and Weber (2013) have found that, indeed, participants tend to be more pro-social in the *Take* frame than in the *Give* frame. Moreover, Krupka and Weber (2013) have shown that the rate of pro-sociality can be predicted by what they called "social appropriateness" of an action. However, the framing effect when passing from the *Take* frame to the *Give* frame has not been replicated by several other works (Dreber et al., 2013; Grossman & Eckel, 2015; Halvorsen, 2015; Hauge et al., 2016; Goerg et al., 2017). This mixed results thus left open the question about whether framing effects actually exist in the Dictator game and whether they can be actually explained in terms of "social appropriateness" or similar constructs. Our work sheds light on this topic. We have indeed found that the words used to describe the available actions in an extreme Dictator game can alter people's behavior, and that, somewhat in line with Krupka and Weber (2013), that the rate of pro-sociality can be predicted by the moral judgments associated to the available actions. However, in line with (Dreber et al., 2013; Grossman & Eckel, 2015; Halvorsen, 2015; Hauge et al., 2016; Goerg et al., 2017) we have found that the *Take* frame does not give rise to a rate of pro-sociality significantly higher than the *Give* frame.

The fact that participants care about doing what they think is the morally right thing has been shown in several contexts in the last years. For example, empirical evidence has been presented to support the hypotheses that (i) altruistic behavior in the Dictator game and cooperative behavior in the prisoner's dilemma are partly driven by a desire to do the "right thing" (Capraro & Rand, 2018; Tappin & Capraro, 2018), (ii) differences in behavior in the Dictator game in the "give" frame vs the "take" frame are driven by a change in people's perception about what is the appropriate thing to do (Krupka & Weber, 2013), (iii) a change in the perception of what is the morally wrong thing to do can generate framing effects among Ultimatum game responders (Eriksson et al., 2017), (iv) moral reminders increase pro-social behavior in both the Dictator game and the prisoner's dilemma (Brañas-Garza, 2007; Capraro et al., 2017; Dal Bó & Dal Bó, 2014). A number of theoretical models have also been introduced to formalize people's tendency towards doing the right thing (Alger & Weibull, 2013; Brekke et al., 2003; DellaVigna et al., 2012; Huck et al., 2012; Kessler & Leider, 2012; Kimbrough & Vostroknutov, 2016; Krupka & Weber, 2013; Lazear et al., 2012; Levitt & List, 2007). Our work adds to this literature by showing that small changes in the language used to describe the available actions can affect the moral judgments associated to those actions, which ultimately affect participants' decisions.

Related to our work is also the game-theoretical literature on language-based games. In fact, the idea that the language used to describe the available strategies can affect people's behavior has been supported also by game theorists. Bjorndahl, Halpern and Pass (2013) have argued that outcome-based preferences do not suffice to explain some human interactions, which are instead best understood by defining the utility function on the underlying language used to describe the game. Motivated by this observation, they have defined the class of language-based games and have studied a generalization of Nash equilibrium and rationalizable strategies on these games.

Of course, our results should be interpreted within the natural limitations of our experiment: an extreme variant of the Dictator Game with $0.50 at stake. For example, although previous research has found very little evidence of stake effects in several games involving pro-sociality when stakes are not too high (Forsythe et al., 1994; Carpenter et al., 2005; Johansson-Stenman et al., 2005; Brañas-Garza et al., 2018; Larney et al., 2019), other studies have found evidence that pro-sociality decreases at very high stakes (Carpenter et al., 2005; Andersen et al., 2011). Along these lines, it might be possible that, when facing very high stakes, participants' behavior becomes less influenced by the words being used to describe the available actions. Understanding the boundary conditions of the effect of the words used to describe the available actions on participants' behavior is a primary direction for future research. Moreover, our results are silent about the channel through which words affect moral judgments. One possibility is that the words generate a continuous endowment effect (Reb & Connolly, 2007) which ultimately affects participants' choices. This is certainly a topic that needs to be thoroughly investigated.

In conclusion, our data show that the words used to describe the available actions can affect people's decisions in extreme dictator games. Future research should explore the boundary conditions of this effect and the conceptual channels through which words affect moral judgments.

References

Alger, I. & Weibull, J. W. (2013). Homo moralis—preference evolution under incomplete information and assortative matching. *Econometrica*, 81(6), 2269–2302.

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., & List, J. A. (2011). Stakes matter in ultimatum games. *American Economic Review*, 101(7), 3427–39.

Arechar, A. A., Gächter, S., & Molleman, L. (2018). Conducting interactive experiments online. *Experimental Economics*, 21(1), 99–131.

Baccianella, S., Esuli, A., & Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner,

& D. Tapias (Eds.), *LREC*: European Language Resources Association.

Bardsley, N. (2008). Dictator game giving: Altruism or artefact? *Experimental Economics*, 11(2), 122–133.

Bjorndahl, A., Halpern, J. Y., & Pass, R. (2013). Language-based games. In *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013* (pp. 2967–2971).

Bolton, G. E. & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1), 166–193.

Brañas-Garza, P. (2007). Promoting helping behavior with framing in dictator games. *Journal of Economic Psychology*, 28(4), 477–486.

Brañas-Garza, P., Capraro, V., & Rascon-Ramirez, E. (2018). Gender differences in altruism on mechanical turk: Expectations and actual behaviour. *Economics Letters*, 170, 19–23.

Brekke, K. A., Kverndokk, S., & Nyborg, K. (2003). An economic model of moral motivation. *Journal of Public Economics*, 87(9-10), 1967–1983.

Cappelen, A. W., Nielsen, U. H., Sørensen, E. Ø., Tungodden, B., & Tyran, J.-R. (2013). Give and take in dictator games. *Economics Letters*, 118(2), 280–283.

Capraro, V., Jagfeld, G., Klein, R., Mul, M., & van de Pol, I. (2017). What's the right thing to do? Increasing pro-sociality with simple moral nudges. *arXiv preprint arXiv:1711.05492*.

Capraro, V. & Rand, D. G. (2018). Do the right thing: Experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality. *Judgment and Decision Making*, 13(1), 99–111.

Carpenter, J., Verhoogen, E., & Burks, S. (2005). The effect of stakes in distribution experiments. *Economics Letters*, 86(3), 393–398.

Chang, D., Chen, R., & Krupka, E. (2017). Rhetoric matters: A social identity explanation for the anomaly of framing. *University of Michigan Workin Paper*.

Charness, G. & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3), 817–869.

Dal Bó, E. & Dal Bó, P. (2014). "Do the right thing:" The effects of moral suasion on cooperation. *Journal of Public Economics*, 117, 28–38.

DellaVigna, S., List, J. A., & Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *The Quarterly Journal of Economics*, 127(1), 1–56.

Dreber, A., Ellingsen, T., Johannesson, M., & Rand, D. G. (2013). Do people care about social context? Framing effects in dictator games. *Experimental Economics*, 16(3), 349–371.

Eckel, C. C. & Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and economic behavior*, 16(2), 181–191.

Eichenberger, R. & Oberholzer-Gee, F. (1998). Rational moralists: The role of fairness in democratic economic politics. *Public Choice*, 94(1-2), 191–210.

Ellingsen, T., Johannesson, M., Mollerstrom, J., & Munkhammar, S. (2012). Social framing effects: Preferences or beliefs? *Games and Economic Behavior*, 76(1), 117–130.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4), 583–610.

Engelmann, D. & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American Economic Review*, 94(4), 857–869.

Eriksson, K., Strimling, P., Andersson, P. A., & Lindholm, T. (2017). Costly punishment in the ultimatum game evokes moral concern, in particular when framed as payoff reduction. *Journal of Experimental Social Psychology*, 69, 59–64.

Esuli, A. & Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06* (pp. 417–422).

Fehr, E. & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3), 817–868.

Fehr, E. & Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, 1, 615–691.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3), 347–369.

Goerg, S. J., Rand, D. G., & Walkowitz, G. (2017). Framing effects in the prisoner's dilemma but not in the dictator game. *Available at SSRN: https://ssrn.com/abstract=2912982 or http://dx.doi.org/10.2139/ssrn.2912982*.

Grossman, P. J. & Eckel, C. C. (2015). Giving versus taking for a cause. *Economics Letters*, 132, 28–30.

Halvorsen, T. U. (2015). Are dictators loss averse? *Rationality and Society*, 27(4), 469–491.

Hauge, K. E., Brekke, K. A., Johansson, L.-O., Johansson-Stenman, O., & Svedsäter, H. (2016). Keeping others in our mind or in our heart? distribution games under cognitive load. *Experimental Economics*, 19(3), 562–576.

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.

Huck, S., Kübler, D., & Weibull, J. (2012). Social norms and economic incentives in firms. *Journal of Economic Behavior & Organization*, 83(2), 173–185.

Johansson-Stenman, O., Mahmud, M., & Martinsson, P. (2005). Does stake size matter in trust games? *Economics Letters*, 88(3), 365–369.

Kay, A. C. & Ross, L. (2003). The perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the prisoner's dilemma. *Journal of Experimental Social Psychology*, 39(6), 634–643.

Kessler, J. B. & Leider, S. (2012). Norms and contracting. *Management Science*, 58(1), 62–77.

Kimbrough, E. O. & Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, 14(3), 608–638.

Krupka, E. L. & Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3), 495–524.

Larney, A., Rotella, A., & Barclay, P. (2019). Stake size effects in ultimatum game and dictator game offers: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 151, 61–72.

Lazear, E. P., Malmendier, U., & Weber, R. A. (2012). Sorting in experiments with application to social preferences. *American Economic Journal: Applied Economics*, 4(1), 136–63.

Levitt, S. D. & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2), 153–174.

Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and social psychology bulletin*, 30(9), 1175–1185.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political economy*, 115(3), 482–493.

Liu, J., Cao, Y., Lin, C.-Y., Huang, Y., & Zhou, M. (2007). Low-quality product review detection in opinion summarization. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.

McGlohon, M., Glance, N., & Reiter, Z. (2010). Star quality: Aggregating reviews to rank products and merchants. In *Proceedings of Fourth International Conference on Weblogs and Social Media (ICWSM)*.

Pang, B. & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics* (pp. 271).: Association for Computational Linguistics.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79–86).: Association for Computational Linguistics.

Paolacci, G. & Chandler, J. (2014). Inside the turk: Under-standing mechanical turk as a participant pool. *Current Directions in Psychological Science*, 23(3), 184–188.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making*, (pp. 411–419).

Rand, D. G. (2012). The promise of mechanical turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299, 172–179.

Rand, D. G., Brescoll, V. L., Everett, J. A., Caparo, V., & Barcelo, H. (2016). Social heuristics and social roles: Intuition favors altruism for women but not for men. *Journal of Experimental Psychology: General*, 145(4), 389.

Reb, J. & Connolly, T. (2007). Possession, feelings of ownership, and the endowment effect. *Judgment and Decision Making*, 2(2), 107.

Swope, K., Cadigan, J., Schmitt, P., & Shupp, R. (2008). Social position and distributive justice: Experimental evidence. *Southern Economic Journal*, (pp. 811–818).

Tappin, B. M. & Caparo, V. (2018). Doing good vs. avoiding bad in prosocial choice: A refined test and extension of the morality preference hypothesis. *Journal of Experimental Social Psychology*, 79, 64–70.

Vanzo, A., Croce, D., & Basili, R. (2014). A context-based model for sentiment analysis in twitter. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 2345–2354).

Zhang, L. & Ortmann, A. (2013). On the interpretation of giving, taking, and destruction in dictator games and joy-of-destruction games. *Preprint available at SSRN: abstract_id=2190240*.

Appendix A: Experimental Instructions of Study 1

Steal vs Don't steal condition

There are 50 cents available. You have to choose between two possible actions:

- STEAL FROM THE OTHER PARTICIPANT: In which case, you get the 50 cents and the other participant gets 0 cents;

- DON'T STEAL FROM THE OTHER PARTICIPANT: In which case, you get 0 cents and the other participant gets 50 cents.

The other participant has no choice and will be paid according to your decision. No deception is used. You and the other participant will be paid according to your decision.

Here are some questions to ascertain that you understand the rules. Remember that you have to answer all of these

questions correctly in order to get the completion code. If you fail any of them, the survey will automatically end and you will not get any payment.[5]

- What choices should YOU make in order to maximize YOUR gain? [*Available answers: Steal from the other participant – Don't steal from the other participant.*]

- What choice should YOU make in order to maximize the OTHER PARTICIPANT's gain? [*Available answers: Steal from the other participant – Don't steal from the other participant.*]

Congratulations, you passed all comprehension questions. It is now time to make your choice.

- What is your choice? [*Available options: Steal from the other participant (50 cents for you, 0 cents for the other participant) / Don't steal from the other participant (0 cents for you, 50 cents for the other participants)*]

Don't steal vs Steal condition

Identical to the previous one, with the only difference that the word "Steal" was replaced by "Don't Steal", and the words "Don't steal" were replaced by "Steal". Payoffs were changed accordingly.[6]

Appendix B: Experimental Instructions of Study 2

[The following questions were asked in random order]

Steal screen

Imagine that there are 50 cents available and that you have to choose between two possible actions:

Steal from the other participant: In which case, you get the 50 cents and the other participant gets 0 cents

Don't steal from the other participant: In which case, you get 0 cents and the other participant gets 50 cents.

Having this situation in mind, please answer the following questions [presented in random order]:

From a moral point of view, how would you judge the choice: to steal? [Available answers: Extremely wrong / Somewhat wrong / Neutral / Somewhat right / Extremely right]

From a moral point of view, how would you judge the choice: not to steal? [Available answers: Extremely wrong

/ Somewhat wrong / Neutral / Somewhat right / Extremely right]

Take screen

Imagine that there are 50 cents available and that you have to choose between two possible actions:

Take from the other participant: In which case, you get the 50 cents and the other participant gets 0 cents

Don't take from the other participant: In which case, you get 0 cents and the other participant gets 50 cents.

Having this situation in mind, please answer the following questions [presented in random order]:

From a moral point of view, how would you judge the choice: to take? [Available answers: Extremely wrong / Somewhat wrong / Neutral / Somewhat right / Extremely right]

From a moral point of view, how would you judge the choice: not to take? [Available answers: Extremely wrong / Somewhat wrong / Neutral / Somewhat right / Extremely right]

Demand screen

Imagine that there are 50 cents available and that you have to choose between two possible actions:

Demand from the other participant: In which case, you get the 50 cents and the other participant gets 0 cents

Don't demand from the other participant: In which case, you get 0 cents and the other participant gets 50 cents.

Having this situation in mind, please answer the following questions [presented in random order]:

From a moral point of view, how would you judge the choice: to demand? [Available answers: Extremely wrong / Somewhat wrong / Neutral / Somewhat right / Extremely right]

From a moral point of view, how would you judge the choice: not to demand? [Available answers: Extremely wrong / Somewhat wrong / Neutral / Somewhat right / Extremely right]

Give screen

Imagine that there are 50 cents available and that you have to choose between two possible actions:

Give to the other participant: In which case, you get the 0 cents and the other participant gets 50 cents

Don't give to the other participant: In which case, you get 50 cents and the other participant gets 0 cents.

Having this situation in mind, please answer the following questions [presented in random order]:

From a moral point of view, how would you judge the choice: to give? [Available answers: Extremely wrong /

---

[5]A skip logic in the survey eliminated from the survey automatically all participants providing the wrong answer

[6]An identical scheme has been adopted for the *Take vs Don't take*, *Don't take vs take*, *Demand vs Don't demand*, *Don't demand vs Demand*, *Give vs Don't give*, *Don't give vs Give*, *Donate vs Don't donate*, *Don't donate vs Donate*, *Boost vs Don't boost*, and *Don't boost vs Boost* conditions.

Somewhat wrong / Neutral / Somewhat right / Extremely right]

From a moral point of view, how would you judge the choice: not to give? [Available answers: Extremely wrong / Somewhat wrong / Neutral / Somewhat right / Extremely right]

Donate screen

Imagine that there are 50 cents available and that you have to choose between two possible actions:

Donate to the other participant: In which case, you get the 0 cents and the other participant gets 50 cents

Don't donate to the other participant: In which case, you get 50 cents and the other participant gets 0 cents.

Having this situation in mind, please answer the following questions [presented in random order]:

From a moral point of view, how would you judge the choice: to donate? [Available answers: Extremely wrong / Somewhat wrong / Neutral / Somewhat right / Extremely right]

From a moral point of view, how would you judge the choice: not to donate? [Available answers: Extremely wrong / Somewhat wrong / Neutral / Somewhat right / Extremely right]

Boost screen

Imagine that there are 50 cents available and that you have to choose between two possible actions:

Boost the other participant: In which case, you get the 0 cents and the other participant gets 50 cents

Don't boost the other participant: In which case, you get 50 cents and the other participant gets 0 cents.

Having this situation in mind, please answer the following questions [presented in random order]:

From a moral point of view, how would you judge the choice: to boost? [Available answers: Extremely wrong / Somewhat wrong / Neutral / Somewhat right / Extremely right]

From a moral point of view, how would you judge the choice: not to boost? [Available answers: Extremely wrong / Somewhat wrong / Neutral / Somewhat right / Extremely right]