Games and Economic Behavior 135 (2022) 241–270

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

The asymmetric effect of narratives on prosocial behavior

Adrian Hillenbrand[a,b], Eugenio Verrina[c,*,1]

[a] Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany
[b] ZEW – Leibniz Centre for European Economic Research, 68161 Mannheim, Germany
[c] Groupe d'Analyse et de Théorie Economique (GATE), UMR5824, Univ Lyon, CNRS, F-69130 Ecully, France

ARTICLE INFO

Article history:
Received 22 October 2020
Available online 30 June 2022

JEL classification:
C91
D63
D64
D83
D91

Keywords:
Narrative
Prosocial behavior
Motivated reasoning
Dictator game
SVO

ABSTRACT

We study how positive narratives (stories in favor of a prosocial action) and negative narratives (stories in favor of a selfish action) influence prosocial behavior in a series of lab and online experiments with more than 1500 subjects. We find that, both positive and negative narratives are effective at changing how actions are perceived. However, while positive narratives increase prosocial behavior, negative narratives do not move aggregate behavior and — if anything — lead to slightly more prosocial behavior. Our results indicate that this may be due to the fact that when following a negative narrative an individual is viewed as influenceable — something that appears to be undesirable. Taken together, our study suggests that positive and negative narratives are not just the flip sides of the same coin.

1. Introduction

Imagine that for some days you have seen a beggar on your way to work. As you pass by today, you reach into your pocket to get some change. While doing so, you remember what a colleague told you the day before. He stated that most of these people are not really needy, but have simply chosen to live this way. Besides, giving them money does not really help them as they will become dependent on charity. Now imagine your colleague telling you instead that hard economic conditions are causing more and more layoffs and hardship. Giving something can help these people to get another chance. Will you give something to the beggar after recalling one of the two stories? Will you give him more or less than what you had picked from your pocket in the beginning? Will you react differently based on your first tendency to give or not to give something?

[*] Corresponding author.
E-mail addresses: adrian.hillenbrand@zew.de (A. Hillenbrand), verrina@gate.cnrs.fr (E. Verrina).

[1] We thank Zvonimir Bašić, Anna Baumert, Claudia Cerrone, Christoph Engel, Fabio Galeotti, Peter Schwardmann, Matthias Stefan, Joel van der Weele, Marie Claire Villeval, Roberto Weber, Fabian Winter and researchers at the Max Planck Institute for Research on Collective Goods as well as seminar participants at several conferences and workshops for helpful conversations. Financial support by the Max Planck Society, University of Mannheim, ZEW – Leibniz Centre for European Economic Research, and GATE-LAB Funding Programme is gratefully acknowledged. Eugenio Verrina also acknowledges financial support from the French National Research Agency (ANR, DECISION project, ANR-19-CE26-0019). This experiment falls under the Generalized Approval of Standard Economic Experiment given by the Ethics Council of the Max Planck Society (Application No: 2018_3).

https://doi.org/10.1016/j.geb.2022.06.008
0899-8256/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Theoretical accounts of motivated moral reasoning (Ditto et al., 2009) emphasize people's deep need to justify their moral behavior not only to others, but especially to themselves. From a fully rational standpoint, these justifications could reflect pieces of evidence an individual uses to inform her choice. However, cognitive dissonance theory (Festinger, 1962) indicates how such reasons can often be used beyond that to resolve tensions between beliefs and actions (Akerlof and Dickens, 1982).[2] In our opening illustration, the tension between a self-interested and a prosocial action can be resolved differently, depending on the story one is told or recalls. We will call these rationales or justifications that target the perception of appropriateness of a prosocial behavior or the deservingness of the recipient of such behavior *narratives*. The notion of narratives is deeply grounded in psychological theories (McAdams, 1988; Bruner, 1991), where they serve as tools people use to construct their own account of the world. As such, narratives accompany nearly all our decisions, often playing a decisive role in shaping them. Their relevance for economic outcomes has recently received growing attention. Narratives can help explain fluctuations in markets (Shiller, 2017) and also broader historical phenomena (Akerlof and Snower, 2016). Recent theoretical work by Bénabou et al. (2020) has contributed to the understanding of how narratives[3] affect moral or prosocial behavior. The authors develop a model in which individuals with self and social image concerns produce and consume narratives as signals complementing their actions.[4] Unfortunately, naturally occurring data do not allow to isolate the effect of these moral arguments, since they often are bundled together with other types of information. This poses serious challenges in getting at the causal effect of narratives as rationales in favor of a certain behavior.

In this paper, we test how narratives affect prosocial[5] behavior by combining laboratory and online experiments. In our main experiment, subjects play a dictator game in which they decide how to share a given amount of money with another anonymous participant. In two treatment conditions, we look at how positive and negative narratives that people use to justify their own behavior influence the choice of others. We follow Bénabou et al. (2020) and define positive narratives as arguments endorsing moral or prosocial behavior. Negative narratives, on the other hand, are arguments justifying immoral or selfish behavior. Narratives in the NEGATIVE condition are arguments in favor of the selfish action, i.e., giving nothing to the other participant, while narratives in the POSITIVE condition are reasons in favor of the prosocial action, i.e., splitting the amount of money equally.[6] The narratives are constructed by capitalizing on arguments subjects use in previous experimental sessions for justifying their own choice. This confers greater internal validity to our experimental design and allows us to systematically study the effect of the content of narratives, i.e., their appeal to the selfish or the prosocial action.

We compare our two treatments to a BASELINE condition with no narratives. Importantly, we keep empirical expectations across all our conditions constant by showing subjects a distribution of choices made in similar dictator game experiments. This ensures that our treatment manipulations do not carry any valuable empirical information about the relative frequency of choices. We thus isolate the causal effect of narratives as providing or highlighting reasons for either the selfish or the prosocial action.

A key feature of our design is that it allows us to explore how heterogeneous prosocial concerns interact with positive and negative narratives by using subjects' Social Value Orientation (SVO). Heterogeneity in this dimension plays an essential role in theories explaining prosocial behavior (see, e.g., Bénabou and Tirole, 2006) and recent empirical evidence confirms that individuals' prosocial inclination greatly vary (Falk et al., 2018).

We provide a theoretical framework to illustrate how externally supplied narratives influence giving of types with different prosocial orientations and derive simple hypotheses to benchmark our experimental results. In our framework social types are determined by an individual's perception about the deservingness of the recipient and the appropriateness of giving. Narratives are arguments targeting these perceptions. According to our predictions, positive narratives should increase aggregate giving, while negative narratives should decrease it. The effect should go in the same direction for all social[7] types and should be stronger for prosocial types who receive a negative narrative and selfish types who receive a positive narrative.

[2] Epley and Gilovich (2016) make a very similar point in their discussion of the mechanics behind motivated reasoning in general.

[3] Bénabou et al. (2020) also discuss "imperatives", i.e., statements issued by a moral authority dictating to follow a given behavior, as an alternative way to convey moral arguments. The authors present a model, in which a principal who cares about the welfare of an agent can choose to send her either a narrative or an imperative. We focus on settings in which no such authority exists or in which she does not have enough persuasive power to issue an imperative.

[4] Foerster and van der Weele (2021) work out a similar model where a sender with social image concerns can send signals about the social returns to an investment in a public good to a receiver in a pre-play communication phase. Their model generates a set of predictions about the use of the signals which are comparable with Bénabou et al. (2020) for what concerns the focus of this paper. The authors also test their model in an experiment and find support for it.

[5] We focus on prosocial behavior as an important component of moral behavior. As opposed to prosocial behavior, we equate immoral behavior to selfish behavior.

[6] Krupka and Weber (2013) provide compelling empirical evidence that the equal split is indeed considered to be the most socially appropriate behavior in the dictator game and we replicate their findings (see Section 5). In this sense, what we label as the prosocial action would correspond to the social norm, while what we call the selfish action would be the strongest possible deviation from the social norm or the most inappropriate behavior. As hinted in our behavioral predictions (see Section 3.3), our hypotheses also hold in a social norms framework.

[7] We use the term 'social' types to indicate all individuals with different prosocial orientations and the terms 'prosocial' (or prosocials) and 'selfish' to refer to individuals with high or low prosocial concerns.

Our main results are based on a laboratory experiment with 282 subjects and an extensive robustness check with 689 subjects, which was run online.[8] Our results show a consistent positive effect of positive narratives on giving across the two experiments, in line with our predictions. In particular, positive narratives increase the probability to give 5 and decrease the probability to give 0 for selfish types. On the contrary, the effect of negative narratives runs counter our predictions. In the lab experiment, we observe a differential effect on different social types. Prosocial types decrease their giving, while selfish types increase their giving. In the online experiment, negative narratives do not have a significant effect on the level of giving for any type.

We shed some more light on the mechanisms through which narratives work and provide a more encompassing picture by running four additional experiments with a total of 591 subjects. First, we test whether, as our theoretical framework assumes, narratives shift the appropriateness of prosocial and selfish behavior. We find that positive (negative) narratives successfully increase (decrease) the appropriateness of sharing the pie equally and decrease (increase) the appropriateness of keeping everything for oneself. Interestingly, the latter effect is stronger, i.e., while increasing the appropriateness of the action they are advocating for, narratives are even more successful in undermining the appropriateness of the opposite action. Second, we investigate the effect of narratives on image concerns.[9] An important feature of narratives is that they affect how the action of the decision maker is perceived, as argued by Bénabou et al. (2020).[10] Acting selfishly in the presence of a positive narrative sends a clear signal that the decision maker is selfish. In the presence of a negative narrative, on the contrary, the signal is weaker, as selfish decision makers can pool with prosocial ones who follow the narrative. We find that narratives work as intended: individuals who act selfishly in the presence of positive (negative) narratives are viewed as less (more) prosocial. However, this also comes at a cost, as individuals who follow narratives are seen as more influenceable, which is an undesirable trait (see Bursztyn et al., 2020). To get a complete picture of image concerns, we also check their importance for decision makers themselves. Coherently with the findings above, we show that positive (negative) narratives raise (diminish) the extent to which decision makers care about how prosocial their actions are perceived to be. At the same time, their concerns about how influenceable they are perceived to be are unaffected by narratives. Finally, we provide a closer look at how narratives are used by different social types to justify their behavior and whether this changes across treatment conditions.

This evidence helps explain the fact that, although both positive and negative narratives effectively change how actions are perceived, only the former influence behavior. We argue that this happens because concerns about being perceived as influenceable increase relatively to concerns about other traits in the presence of negative narratives. Hence, if decision makers care enough about not being perceived as influenceable negative narratives might become ineffective. We also discuss our findings in relation to the literature on the ‘focusing’ effect.

Our work sheds some first light on how narratives in the realm of prosocial behavior work. From a practical standpoint, our results suggest that making narratives circulate can have positive effects, even if some advocate for selfish behavior. Additionally, the evidence we present suggests that the mechanisms through which positive and negative narratives work are subtly different and that the two are not just the flip sides of each other.

2. Related literature

Our work resonates with the growing interest in the role played by narratives (Bénabou et al., 2020; Foerster and van der Weele, 2021; Shiller, 2017; Akerlof and Snower, 2016; Andre et al., 2021) and, more generally, in the role motivated reasoning plays in shaping economic interactions (Karlsson et al., 2004; Epley and Gilovich, 2016; Bénabou and Tirole, 2016; Golman et al., 2016; Gino et al., 2016; Carlson et al., 2020; Saucet and Villeval, 2019). Our work is also closely linked to experimental studies on phenomena of so-called moral wiggle room (Dana et al., 2007; Larson and Capra, 2009; Matthey and Regner, 2011; van der Weele et al., 2014; Feiler, 2014) and to the wider literature investigating self-serving judgments of fairness or morality (Konow, 2000; Hamman et al., 2010; Shalvi et al., 2011; Wiltermuth, 2011; Rodriguez-Lara and Moreno-Garrido, 2012; Bicchieri and Mercier, 2013; Gino et al., 2013; Shalvi et al., 2015; Exley, 2015; Collins et al., 2018) and self-serving beliefs (Haisley and Weber, 2010; Chance et al., 2011). The main result one can draw from this huge body of evidence is that prosocial behavior is sensitive to the specific context in which choices take place, and that people often tweak the evidence in their favor in conscious and unconscious ways. Our work contributes to this growing literature by providing evidence on how people react to externally provided narratives and by analyzing how heterogeneity in prosocial concerns affects behavior in this context.

Andreoni and Rao (2011) study a setting in which receivers and dictators in a dictator game can communicate with each other. They find that giving increases whenever receivers can say something. Whereas, if only dictators have the word, giving decreases. We investigate a setting in which dictators are exposed to arguments coming from other dictators, who behaved

[8] The main difference between the two is that in the lab the dictator game is played under role-uncertainty, whereas online the roles are assigned from the beginning.

[9] We here refer to both social and self-image concerns, as we believe that the mechanisms described in the following could apply to both. This is in line with previous literature that models the two as parallel reputational motives (Bénabou and Tirole, 2006; Bénabou et al., 2020).

[10] In contrast to Bénabou et al. (2020), we consider exogenous narratives, i.e., dictators receive a narrative from an external source and cannot decide whether to disclose it or not. Our results regarding the effect of narratives on how actions are perceived are in line with the predictions for the case in which narratives are disclosed by the dictators themselves (see Section 5).

either prosocially or selfishly. People are constantly exposed to such arguments both in their professional and private life. We systematically study their effect on prosocial behavior. Similarly, Mohlin and Johannesson (2008) find a positive effect of one-way communication from the receiver to the dictator and also from past receivers to dictators. Differently from these and other studies of communication in economic games (see, e.g., Bohnet, 1999; Charness and Dufwenberg, 2006), we do not look at the effect of communication between parties involved in the game. Instead, we analyze the effect of justifications or rationales, i.e., narratives, that individuals provide for their own choice on the behavior of other individuals facing the same decision.

Other work has looked at how social information (Krupka and Weber, 2009; Gino et al., 2009; Cappelen et al., 2013, 2017) influences prosocial behavior. We hold these channels constant and explicitly provide reasons, or narratives, for a certain action. Thus, our setup allows us to study the causal effect of the *content* (positive or negative) of a narrative on prosocial behavior. In this sense, narratives are conceptually related to framing effects (Andreoni, 1995; Brañas-Garza, 2007; Dreber et al., 2013).

This links our work to studies investigating the effect of moral reminders or recommendations on behavior (see, e.g., Galbiati and Vertova (2008) on obligations and Croson and Marks (2001) on recommendations, both in the public-good game, or Mazar et al. (2008) in the context of lying; further work by Bott et al. (2019) uses moral appeals in letters to tax payers). Most closely related to our paper is an experiment by Dal Bó and Dal Bó (2014), who look at the effect of moral suasion in the form of arguments issued by an authority,[11] i.e., the experimenter, in favor of the socially optimal contribution in a voluntary contribution game. In contrast to them, we look at a non-strategic setting where narratives can only affect preferences and cannot work as coordination devices. Moreover, our messages do not come directly from the experimenter, but are naturally occurring reasons subjects in previous sessions provide for their choices. Last but not least, measuring prosocial concerns allows us to look at heterogeneous effects on different social types and to test the effect of what we call negative narratives more thoroughly.[12]

To achieve this goal, we use the SVO slider measure by Murphy et al. (2011) to measure social types. The SVO measure is a reliable and carefully constructed measure that has been widely used in both psychology and economics to assess heterogeneity in individual motives in social and moral dilemmas (see Balliet et al., 2009, for a meta-study on SVO and cooperation in social dilemmas), e.g., in the public-good game (see e.g. Offerman et al., 1996). Other studies find that individuals scoring differently on the SVO measure exhibit different behavior also in other realms, such as inter-group conflict (Weisel and Zultan, 2016), in vaccine-related behavior (Böhm et al., 2016), and in pay what you want settings (Krämer et al., 2017). Grossman and Van Der Weele (2017) study a setting where people can remain ignorant about harmful consequences of their actions, and find that the SVO measure confirms the sorting predictions of their model. In line with previous studies, we are interested in how heterogeneous prosocial concerns interact with our treatment manipulations.

3. Experimental design

In this section, we describe the experimental design of our main experiment.[13] Our design consists of two separate parts: a type elicitation experiment and a dictator game experiment. The type elicitation experiment was conducted always before the dictator game experiment. The dictator game experiment was implemented in a between-subjects design with a Baseline and two treatment conditions (Positive and Negative), which differ only in the content of the narratives subjects saw. Below, we discuss the individual parts of the study as well as the differences between the lab and online experiment.

Dictator game. The central part of our design is a simple dictator game (Kahneman et al., 1986). Dictators chose how to divide 10 points between themselves and an anonymous recipient. Crucially, we fixed subjects' empirical expectations about the distribution of giving in the dictator game. This makes sure subjects could not take the narratives in our treatment conditions as signals about the empirical distribution of giving. Subjects in all experimental conditions were presented with a graph showing the distribution of dictator game giving in similar experiments.[14] The graph displays data from Engel (2011) restricted to studies in which 10 units of currency were used. Subjects were told the graph displayed the distribution of choices other subjects had made in similar previous experiments. The figure displays the typical bimodal distribution with modes at 5 and 0 with a sizeable mass in between. While holding empirical beliefs constant across our experimental conditions, the distribution does not clearly emphasize one allocation choice over the other.

Treatments. Participants were randomly allocated to one of three treatment conditions in a between-subjects design. In the Baseline condition, subjects only saw the distribution of dictator game giving described above. In the two treatment conditions, they were additionally shown two comments which subjects in the Baseline condition had used to explain

[11] The moral suasion treatments in Dal Bó and Dal Bó (2014) is very close to the notion of imperatives in Bénabou et al. (2020). In this sense, our study and the one by Dal Bó and Dal Bó (2014) can be understood as testing the effect of narratives and that of imperatives, respectively.

[12] Dal Bó and Dal Bó (2014) find that messages explaining the game-theoretical prediction of zero contribution have no effect on contributions. However, baseline contributions are already quite low when they introduce this manipulation and there is hardly any room for a further decrease to take place.

[13] The additional experiments are described in Section 5. More details can be found in Appendix C.

[14] The graph was displayed on subjects' decisions screen which is displayed in Fig. D.1 for the lab and Fig. D.2 for the online experiment.

their choices.[15] These are our narratives (see Appendix D.1.1 for the lab and Appendix D.2.1 for the online experiment). In the Positive condition, subjects saw two comments in support of the equal split (giving 5 points), while in the Negative condition they saw two comments justifying selfish behavior (giving 0 points). Subjects were (truthfully) told that these were explanations other participants had given for their choices in similar previous experiments. As such these narratives posses great ecological validity for the task at hand. In the next paragraph, we explain how we collected and selected the narratives to devise our treatment conditions.

Narrative selection. At the end of the experiment, subjects were given the opportunity, without any prior notice, to explain the reasoning behind their choice in the dictator game. We used the explanations from the Baseline condition to build the set of narratives subjects saw in the Positive and Negative condition. Independent raters, who were blind to the research question, evaluated the narratives along several dimensions. First, they were asked which was the most likely choice (0,1,2, etc.) the decision maker who wrote the narrative had taken. Then, they evaluated how convincing they perceived the narrative to be (on a 7-point Likert scale). We then selected the most convincing narratives in support of giving 0 points and in support of giving 5 points (using average ratings). This ensures comparability across the two treatment conditions and isolates the effect of the argument (positive or negative) provided in the narrative. More details about the procedure used can be found in Appendix D.1.1 for the lab and Appendix D.2.1 for the online experiment.

Type elicitation. The type elicitation experiment was conducted separately from the dictator game experiment to avoid contamination across the two.[16] The purpose of this part of the experiment was to measure subjects' prosocial concerns. Our main measure of a subject's social type is the SVO slider measure (Murphy et al., 2011). Subjects are confronted with 6 choices where they have to trade off their earnings with those of another subject under different budget constraints. From these choices, the so-called SVO angle is constructed, which represents the relative weight subjects put on the payoff of others compared to their own. Subjects with an SVO angle of 0° care only about their payoff, while those with an SVO angle of 45° weigh their payoff and that of the other subject equally. Types with an SVO angle between $-12.04°$ and $22.45°$ are generally classified as individualists (selfish) and those between $22.45°$ and $57.15°$ as prosocials.[17] For further details on the measure, we refer to Murphy et al. (2011).

The SVO measure has been shown to be a stable and consistent predictor of behavior in different social dilemma settings (see Balliet et al., 2009, for a meta-study). Moreover, high SVO types (prosocials) have been shown to differ from low SVO types (selfish) in their decision-making process (e.g., Fiedler et al., 2013). This makes the SVO measure particularly suitable for capturing heterogeneity in reactions to our narrative manipulation.

3.1. Lab experiment

The main difference between the lab experiment and the online experiment lies in the way in which roles were assigned. All subjects in the lab experiment decided under role uncertainty, i.e., each subject made her choice in the role of the dictator and roles were randomly assigned at the very end of the experiment.[18]

After the dictator game decision, we elicited subjects' feelings (see Appendix D.1.2 for more details and Appendix B.4 for an analysis of these measures). At the end of the experiment, we also elicited a series of psychological measures (see Appendix D.1.2 for more details and Appendix B.3 for an analysis of these measures). The lab experiment was conducted at the DecisionLab of the Max Planck Institute for Research on Collective Goods in Bonn between May and June 2018. The type elicitation experiment was conducted using Qualtrics, while the dictator game experiment was programmed in zTree (Fischbacher, 2007). Subjects were recruited via Orsee (Greiner, 2015). In total, 282 subjects (64% female, average age 24.8 years)[19] took part in the experiment. 96 subjects took part in the Baseline condition, 91 in Positive, and 93 in Negative.[20] Each of the 10 points in the dictator game was worth 1€. All subjects received a show-up fee of 5 €, plus their earnings from the type elicitation experiment (2 € participation fee plus between 0.50 € and 3 € for the SVO slider task) and their earnings from the dictator game. Overall, subjects received an average payment of 14.48 €. More details about the lab experiment can be found in Appendix D.1.

[15] In the lab experiment, we randomly selected these narratives from a pool of 4 positive and 4 negative narratives depending on the treatment condition. Since we found that the specific narratives subjects saw did not influence the results, we selected just two positive and two negative narratives per treatment in the online experiment.

[16] For the lab experiment, the type elicitation was conducted online between seven and four days before the dictator game. For the online experiment, that happened four to three days before the dictator game.

[17] Earnings in this task were determined by forming random pairs of subjects and then randomly selecting one of the 6 choices. In the lab experiment, the choice of one of the two subjects in the pair was randomly implemented. In the online experiment, we implemented the dictator's choice and recipients were participants in subsequent studies (the additional experiments).

[18] Iriberri and Rey-Biel (2011) find that role uncertainty decreases selfish choices compared to when subjects play in their actual role. To the extent to which the decrease is not excessive and does not interact with our treatment manipulations, this does not constitute a problem for our design. Also, in the online experiment we assign roles from the beginning to assess the robustness of our results.

[19] For 74 subjects, this information was not recorded.

[20] For the analysis, we exclude 2 subjects who had not taken part in the type elicitation experiment.

3.2. Online experiment

Differently from the lab experiment, in the online experiment roles were assigned from the beginning. This variation in the design was undertaken to check the robustness of the findings from the lab experiment. To adapt to the online format the experiment was run asynchronously, i.e., dictators took their choice in one study and recipients were recruited and payed out in subsequent studies (the additional experiments).[21] Dictators were aware of this and knew that the recipients would also take part in a study of similar length as them. Note that, although the main design changes are not dramatic, the lab and the online settings differ in many respects, e.g., platform, subject pool, attention of subjects. Hence, this can be seen as a rather demanding robustness check.

At the end of the type elicitation experiment, we elicited a psychological measure of social comparison (see Appendix D.2.2 for more details on the measure used and Appendix B.5 for an analysis of this measure). The online experiment was conducted on Prolific between June and October 2021. Qualtrics was used to program the whole experiment. 689 dictators (49.4% female, average age 31.2 years) took part in the experiment. 224 subjects took part in the Baseline condition, 227 in Positive, and 238 in Negative. Each of the 10 points in the dictator game was worth £0.25. All subjects were paid £0.75 for participation in the type elicitation and £0.75 for the participation in the dictator game experiment, plus their earnings from the dictator game and the type elicitation (between £0.50 and £1). The average total payment was £4. More details about the online experiment as well as information about procedures used to ensure high data quality and an analysis of potential demand effects can be found in Appendix D.2.

3.3. Behavioral predictions

We develop a simple theoretical framework describing how prosocial behavior is influenced by narratives and derive benchmark predictions for the effect of our treatment conditions. Our approach builds on Bénabou et al. (2020), from which we borrow some key notions. While their aim is to study a broad set of phenomena, such as the emergence of narratives and their transmission, we focus on getting a deeper understanding of the potentially heterogeneous effects of positive and negative narratives on different social types.[22] This gives us a self-contained theoretical framework for which we provide an intuitive description below (the full version can be found in Appendix A). We first outline the reasoning leading up to our hypothesis on aggregate behavior, and then further qualify our predictions for heterogeneous social types.

We start with the notion that decision makers are more inclined to act prosocially the more the consequences of their actions benefit others or the public good (see, e.g., Goeree et al., 2002, and the discussion in Bénabou and Tirole, 2006). In turn, this influences the extent to which an action is perceived as appropriate. As the literature on social norms shows, changes in what is perceived as socially appropriate reliably predict changes in behavior across several settings (Krupka and Weber, 2013).[23] Similarly, decision makers care about the deservingness of the recipient(s) of their prosocial action. In distributional choices, decision makers want to avoid giving too much to an undeserving recipient and too little to a deserving recipient (Cappelen et al., 2013). However, the true deservingness of recipients is often unknown in the real world (Cappelen et al., 2018). Likewise, the perception of what is deemed as appropriate is highly flexible and prone to self-serving interpretations (Gino et al., 2016).

Narratives in our setting are arguments targeting these perceptions of deservingness or appropriateness. A positive narrative could, for example, state that the recipient is as deserving as the dictator, because both spent the same time in the experiment or because roles were assigned based on chance. By contrast, a negative narrative might undermine the perceived appropriateness of giving, e.g., by arguing that it is not necessary to give to an anonymous recipient or that everyone else would also behave selfishly, questioning the deservingness of other participants. Importantly, these stories only need to be convincing in the sense of influencing a decision maker's perception of the situation. If positive or negative narratives are indeed successful in changing the perception of the decision maker, they will influence behavior. Our hypothesis on aggregate behavior follows directly.

Hypothesis 1. Positive narratives increase giving, while negative narratives decrease giving.

We now look at how different social types are influenced by negative and positive narratives. As mentioned above, the deservingness of a recipient and the appropriateness of giving are subject to uncertainty. This uncertainty leaves room for diverging perceptions which, in our model, define a decision maker's type.[24] In our setting, we call decision makers who

[21] Recipients were recruited in the same way and with the same restrictions as subjects in the main experiment. They learned that they were the recipients of the dictator game only at the very end of their experiment and after they made all choices.

[22] In the model by Bénabou et al. (2020), types are defined as either moral or immoral. In our setting, we look at a continuum of types, where heterogeneity stems from diverging beliefs about the appropriateness of an action or the deservingness of the recipient of this action.

[23] The main intuitions we derive from our theoretical framework also hold in a social norms environment with heterogeneous beliefs about the appropriateness to follow the norm, as we describe in Appendix A.

[24] We are agnostic about where these different perceptions come from and simply require them to influence behavior. They may be deeply grounded in a decision maker or may have formed through experience, or else a decision maker might self-servingly hold a perception which allows her to act in a certain way.

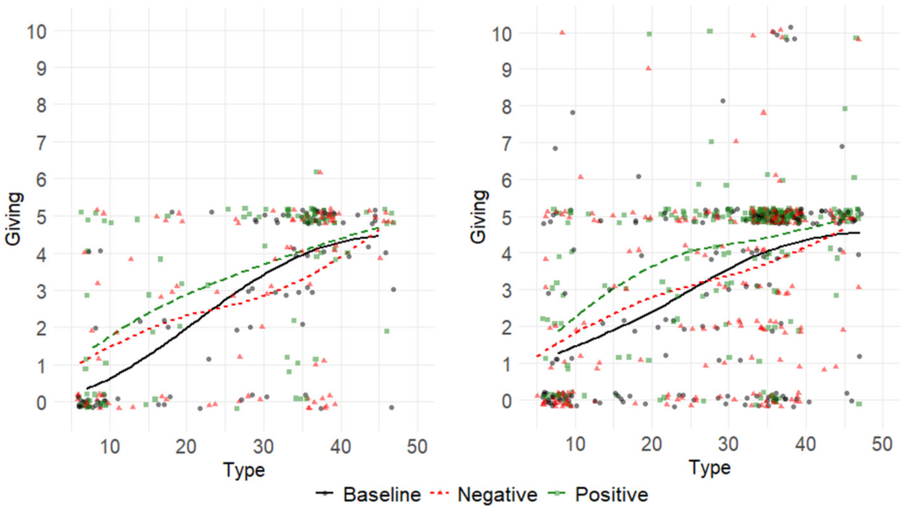


Fig. 1. Giving on SVO, LOESS fitted lines (lab experiment on the left and online experiment on the right). Note: Data points are jittered. Black circles represent observations in the Baseline, green triangles in the Positive and red squares in the Negative treatment. Green, dashed lines represent the Positive treatment, red dotted lines the Negative and the solid black line represents the Baseline. For the ease of visualization, we removed social types below 7° and above 50°, which are rare and not balanced across treatments (5 subjects in the lab and 26 in the online experiment).

initially (i.e., in the absence of a narrative) perceive a recipient to be deserving or giving to be appropriate 'prosocial' types and the ones who believe the opposite 'selfish' types.[25]

Consider a prosocial decision maker who hears a negative narrative undermining her perception of the recipients' deservingness. If, as we assume above, she ascribes some truth to the narrative, her perception about the appropriateness of giving, and hence her behavior, will change and lead her to give less. Importantly, this effect will be greater compared to that of the same negative narrative on a selfish decision maker, who had a lower perception of the recipients' deservingness in the first place. Vice versa, a positive narrative will have a greater effect on a selfish compared to a prosocial decision maker.

Hypothesis 2. Positive narratives have a stronger effect on more selfish types, while negative narratives have a stronger effect on more prosocial types.

4. Results

In this section, we lay down the results of our main experiment. First, we analyze the evidence regarding our main hypotheses. Then, we provide additional insights on the way our treatment conditions influence behavior by looking at whether subjects follow positive or negative narratives.

4.1. Main results

Fig. 1 provides a first overview of our main results in the lab and the online experiment. Subjects in the Baseline condition give on average 2.76 points in the lab and 3.28 points in the online experiment.[26] According to Hypothesis 1, we should observe an increase in average giving in the Positive condition and a decrease in the Negative condition. In the Positive condition, average giving increases to 3.23 in the lab and 3.99 in the online experiment. This constitutes a 17% and 22% increase in the lab and the online experiment, and is thus in line with our first hypothesis. The difference is marginally significant in the lab (rank-sum test, $p = .093$) and highly significant in the online experiment (rank-sum test, $p = .0015$).[27] Average giving in the Negative condition (2.78 points in the lab and 3.20 points online) is virtually identical to average giving in the Baseline condition in both experiments (rank-sum test, $p > 0.1$ for both the lab and online experiment.)

However, as Fig. 1 shows, the aggregate results on giving provide an incomplete picture of the data as different types seem to react differently to narratives. As stated in Hypothesis 2, prosocial types should respond more strongly to the Negative condition and selfish types to the Positive condition. Although the effect should go in the same direction for all

[25] In our experiment, we use the Social Value Orientation to measure these different perceptions. A higher (lower) SVO angle corresponds to a higher (lower) perception of deservingness or appropriateness.

[26] In the Baseline condition of the online experiment, selfish subjects are more generous compared to the lab experiment. This reduces the room in which narratives can operate and can explain lower marginal effects.

[27] All tests are two-sided.

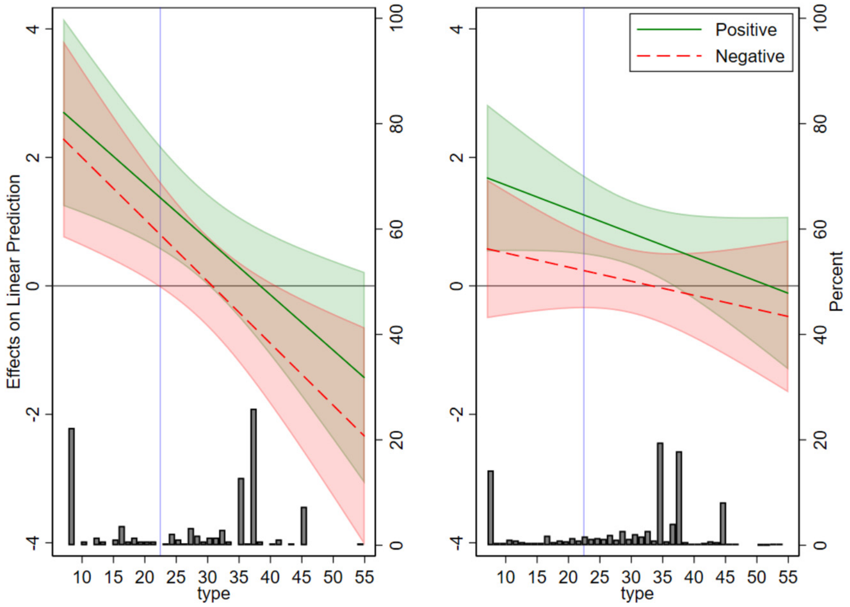


Fig. 2. Marginal effects on types, Tobit (lab experiment on the left and online experiment on the right). Note: In the lower part of the graphs, we plot the pooled distribution of types (as measured by the SVO-angle) over all conditions. For the ease of visualization, types below 7° (3 subjects in the lab, 20 online) are not displayed. Shaded areas show 95% confidence intervals. The vertical blue line divides selfish from prosocial types.

types. To test how different types react to different narratives, we run a Tobit model, as suggested by Engel (2011), with the amount of giving as the dependent variable and treatment dummies, type, and interaction terms between type and treatment dummies as explanatory variables (see Table B.1).[28]

To interpret these results, we plot the estimated marginal effects of our treatment conditions on giving for different types compared to the Baseline in Fig. 2. This enables us to test Hypothesis 2. We start with the Positive condition (green, solid line), where we find a pattern in line with our hypothesis in both the lab and the online experiment. We notice a strong positive effect for selfish types, which fades out for prosocial types. This is confirmed when considering selfish and prosocial types separately using non-parametric tests. For these tests, we restrict our sample to subjects categorized as selfish or prosocial according to (Murphy et al., 2011).[29] Selfish types increase giving by 1.17 points on average in the lab (Baseline vs. Positive, 0.62 vs. 1.79, rank-sum test, $p = 0.0061$) and by 0.58 points in the online experiment (1.83 vs. 2.41, rank-sum test, $p = 0.0936$). For prosocial types there is no significant difference in the lab experiment (3.94 vs. 4.12, rank-sum test, $p = 0.3045$), while the effect is positive and significant in the online experiment (3.95 vs. 4.47, rank-sum test, $p = 0.0195$).

Result 1. Positive narratives increase giving compared to the Baseline condition. This effect is driven mainly by selfish types.

The results of the Negative condition (red, dashed line) run against our hypotheses. In the lab (left graph), we find a positive effect for selfish types and this effect turns negative for more prosocial types. Looking at the two types in isolation with non-parametric tests, we find a sizeable and significant effect for selfish (Baseline vs. Negative, 0.62 vs. 1.83, rank-sum test, $p = 0.0076$), while for the whole set of prosocial types the effect is not significant (3.93 vs 3.33, $p = 0.1588$). In the online experiment (right graph), negative narratives have no significant effect across all types. Non-parametric tests confirm this both for selfish (1.83 vs. 1.91, $p = 0.7799$) and prosocial types (3.95 vs. 3.73, $p = 0.2472$).

Result 2. In the lab, we find a positive effect of negative narratives on selfish types and a (weak) negative effect for more prosocial types compared to the Baseline. In the online experiment, negative narratives do not significantly change behavior.

In Appendix B.2, we provide several robustness checks of these results by adding controls as well as by using other functional forms and regression models. Our results stay largely the same.

[28] The distributions of types do not significantly differ between treatment conditions in the lab or online (Kolmogorov-Smirnov exact test, all $p > 0.05$).

[29] This excludes 5 subjects categorized as altruists and 2 subjects categorized as competitive in the online experiment and 3 subjects categorized as competitive in the lab experiment. Including these subjects in the analysis does not change our results.

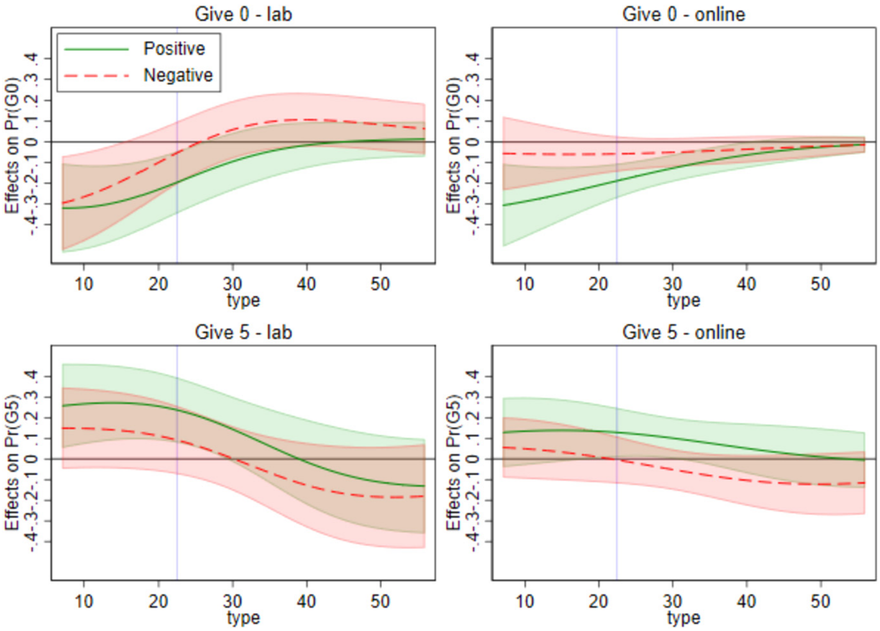


Fig. 3. Marginal effects, Probit (lab experiment on the left and online experiment on the right). Note: The dependent variable is a dummy for giving 0 in the upper two graphs and for giving 5 in the lower two graphs. For the ease of visualization, types below 7° (3 subjects in the lab, 20 online) are not displayed. Shaded areas show 95% confidence intervals. The vertical blue line divides selfish from prosocial types.

4.2. Following the narrative

A natural question is whether narratives lead subjects to adhere to the behavioral prescription contained in them. In other words, did the Positive (Negative) condition lead subjects to give 5 (0) more frequently than in the Baseline? To answer this question, we run a Probit model on the probability of giving either 5 or 0 using the same specification as in our main regression (see Table B.2). The graphs in Fig. 3 show the estimated marginal effects of the Positive and Negative condition on different social types compared to the Baseline in the lab and online experiment.

We first focus on the Positive condition (green, solid lines) and its effect on giving 5 (two lower graphs). Positive narratives increase the probability to give 5 for selfish types, while they do not significantly affect prosocial types. This effect is more pronounced in the lab than in the online experiment. Non-parametric tests confirm this result (Baseline vs. Positive, lab: 3% vs 21%, $p = 0.027$, online: 21% vs 26%, $p = 0.665$, Fisher's exact test). The presence of positive narratives also decreases the probability to give 0 (two upper graphs) for selfish types considerably both in the lab and online experiment. Again, this is confirmed by non-parametric tests (Baseline vs. Positive, lab: 79% vs. 48%, $p = 0.011$, online: 53% vs 31%, $p < 0.001$, Fisher's exact test).

Result 3. The Positive condition increases the probability to give 5 and decreases the probability to give 0 for selfish types.

The effect of the Negative condition (red, dashed lines) on behavior is less pronounced, in line with our main results. Negative narratives decrease the probability of giving 0 (two upper graphs) for selfish types in the lab, but this effect does not transfer to the online experiment. Non-parametric tests also show this difference (Baseline vs. Negative, lab: 79% vs 40%, $p = 0.018$, online: 53% vs. 48%, $p = 0.612$, Fisher's exact test). Thus, if anything, subjects are less likely to follow the negative narrative. We also see an increase in the probability of giving 5 (two lower graphs) after observing a negative narrative for selfish subjects in the lab, but again this effect is not found in the online experiment. Non-parametric tests corroborate this difference (Baseline vs. Negative, lab: 3% vs 20%, $p = 0.027$, online: 21% bs 19%, $p = 0.833$, Fisher's exact test). In both the lab and online experiment, prosocials do not tend to give 5 less often compared to the Baseline. These effects are weak and only marginally significant in the online experiment (Baseline vs. Negative, lab: 60% vs 48%, $p = 0.208$, online: 61% vs 51%, $p = 0.072$, Fisher's exact test).

Result 4. The Negative condition – if anything – decreases the probability to give 0 and decreases the probability to give 5.

5. Additional experiments

We run additional experiments to shed further light on how narratives work. First, we test whether, as our theoretical framework assumes, narratives support or undermine the perceived appropriateness of an action. Second, we investigate the effect of narratives on image concerns. Negative narratives should make it easier to act selfishly by allowing selfish decision makers who would do so regardless to pool with more prosocial types who follow the narrative. Positive narratives should make it harder to act selfishly since decision makers who do so would be clearly identified as selfish. In connection to this, we test whether the presence of narratives might directly influence the extent to which decision makers care about how they are perceived by others. Finally, we also provide a closer look at which narratives are used by different social types to justify behavior and whether this changes across treatment conditions. Below we provide a brief description of the design of each experiment and lay down the results. More details on the experimental design of the additional experiments can be found in Appendix C.

5.1. Appropriateness

To test whether narratives change the appropriateness of selfish and prosocial behavior, we use the method by Krupka and Weber (2013) to obtain the socially shared appropriateness ratings (social norm). Subjects rate the appropriateness of giving 5 or 0 in the presence of positive or negative narratives or with no narrative (in all cases the graph used for the main experiment is displayed to subjects). All results stem from within-subject comparisons (subjects rated all three experimental conditions in random order).

The top graph of Fig. 4 shows the mean appropriateness ratings by treatment condition. First, note that the appropriateness of giving 0 is clearly lower than that of giving 5 across all scenarios. As our theory assumes (see Section 3.3), the presence of positive narratives increases the appropriateness of giving 5 (5.4 on a 6-point likert scale) compared to when there is no narrative (5.0) and decreases the appropriateness of giving 0 (2.4) compared to when there is no narrative (3.3). Similarly, the presence of a negative narrative decreases the appropriateness of giving 5 (4.5) and increases the appropriateness of giving 0 (3.6). Considering individual ratings as observations, all ratings differ significantly between the different scenarios (Wilcoxon signed-rank test, $p < 0.01$ in all pairwise comparisons). Notably, while still different the appropriateness of giving 5 and giving 0 move quite close with a negative narrative (3.6 vs. 4.5). Interestingly, the negative effect on the appropriateness of the action a narrative does not support is stronger than the positive effect on the appropriateness of the action that it supports.

Result 5. Narratives shift the appropriateness of giving 0 and giving 5 in the expected direction. Positive (negative) narratives increase (decrease) the appropriateness of giving 5 and decrease (increase) the appropriateness of giving 0.

5.2. Image concerns

To test whether narratives affect how different choices are perceived, we incentivize subjects to guess the SVO score (measured on a 10-point scale), i.e., the type, of a decision maker in the main experiment depending on whether and which narratives she saw and which action she took. Following a narrative may also make a decision maker appear more influenceable — which is an undesirable trait, as argued by Bursztyn et al. (2020).[30] Hence, we also ask subjects to rate how influenceable a decision maker appeared to them. In a separate experiment, we also look at whether narratives influence how much decision makers themselves care about these traits and about how they are perceived in general.[31]

The two middle graphs of Fig. 4 show the mean ratings of prosociality and influenceability by treatment condition. Our results show that individuals who give 0 are perceived as more prosocial in the Negative condition compared to the Baseline (2.94 vs 2.31, Wilcoxon signrank-test, $p < 0.001$). At the same time, those who give 0 in the Positive condition are seen as less prosocial compared to the Baseline, although this difference is only marginally significant (2.11 vs. 2.31, Wilcoxon signrank-test, $p = 0.0905$). Individuals who give 5, are perceived as not significantly more or less prosocial in the Negative condition compared to the Baseline (7.32 vs 7.54, Wilcoxon signrank-test, $p = 0.631$). While those who give 5 in the Positive condition are seen as more prosocial compared to the Baseline (8 vs. 7.54, Wilcoxon signrank-test, $p = 0.0044$). Note that within treatment conditions one is always perceived as more prosocial when giving 5 compared to giving 0 (Wilcoxon signrank-test, all $p < 0.001$). The same results extend to measures of trustworthiness and likability (see Appendix C).

Result 6. Narratives shift the way others perceive a decision maker's actions in the expected direction. Decision makers who give 0 in the presence of positive (negative) narratives are viewed as less (more) prosocial.

Regarding how influenceable decision makers are perceived to be, we find that individuals who give 0 in the Negative condition, i.e., those who follow the narrative, are seen as more influenceable compared to those that give 0 in the Baseline

[30] In this study, individuals may be considered influenceable because they believe the conclusions of a study despite its methodological flaws and critiques.

[31] Subjects in this experiment received the same information as dictators in the main experiment, but did not take an active decision. This way we avoid answers from being self-servingly influenced by choices. We did not ask these questions to subjects in our main experiment in order not to contaminate their giving decision.

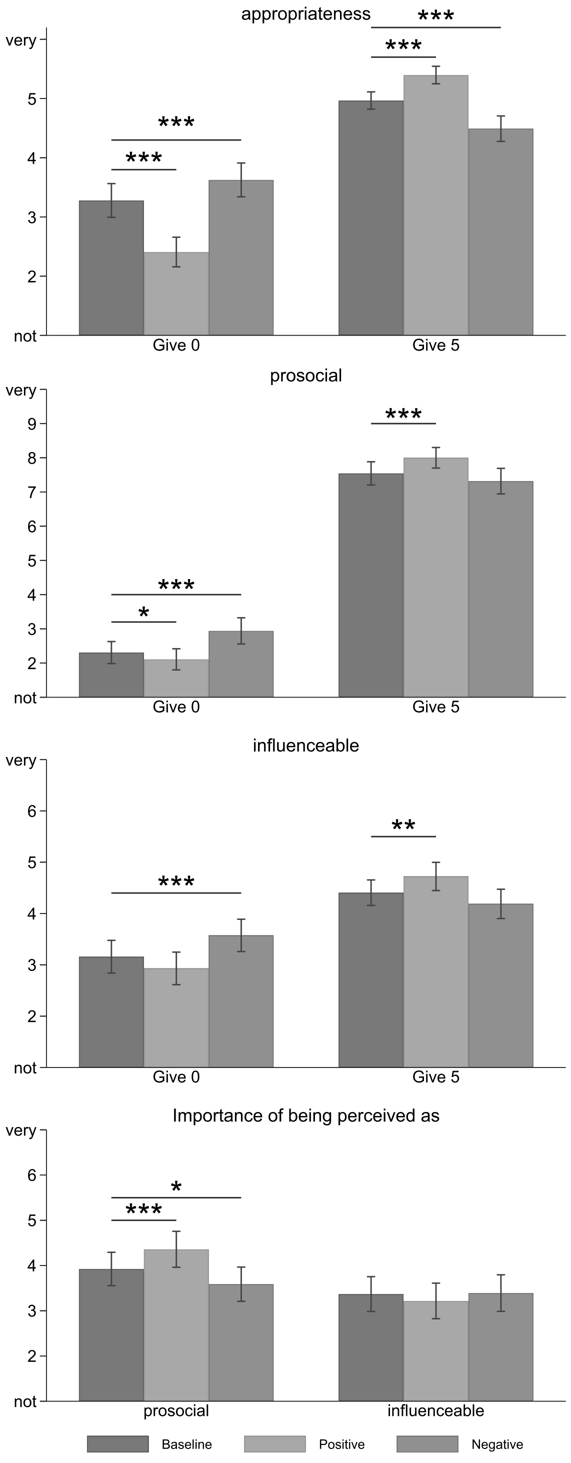


Fig. 4. Appropriateness ratings and image concerns. Note: The top three graphs show mean ratings of appropriateness, prosociality and influenceability for giving 0 and giving 5 across the three different scenarios when no, positive, or negative narratives are present. The bottom graph shows the importance subjects assign to being perceived as prosocial and not influenceable in the three different scenarios. Error bars show 95% confidence intervals. The lines show significant comparisons to the Baseline: * $p < .10$, ** $p < .05$, *** $p < .01$.

(3.57 vs. 3.16, Wilcoxon signrank-test, $p = 0.0018$). Similarly, those who give 5 in the Positive condition are also seen as more influenceable compared to those who give 5 in the Baseline (4.72 vs. 4.41, Wilcoxon signrank-test, $p = 0.0214$).[32]

Result 7. Individuals who follow narratives are seen as more influenceable.

Last we report how much decision makers themselves care about the aspects mentioned above. First, we explicitly ask them how much importance they assign to how their actions are perceived in general. We find that the Positive condition increases this concern compared to the Baseline (4.16 vs. 3.80, Wilcoxon signrank-test, $p = 0.0061$), while the Negative condition decreases it (3.29 vs. 3.80, Wilcoxon signrank-test, $p = 0.0069$). We also focus more specifically on the extent to which decision makers care about the individual traits mentioned above (see bottom graph of Fig. 4). In line with the answer to the general questions, positive narratives increase concerns related to how prosocial someone is perceived to be (4.36 vs. 3.92, Wilcoxon signrank-test, $p = 0.0029$). Negative narratives decrease these concerns, although the effect is only marginally significant (3.59 vs. 3.92, Wilcoxon signrank-test, $p = 0.0680$). The same pattern is present for trustworthiness and likeability, although not always significant (see Appendix C.2.1). Concerns related to how influenceable one is perceived to be are not affected by the presence of either narrative. Hence, the relative importance of influenceability increases (decreases) with negative (positive) narratives.

Result 8. Positive (negative) narratives raise (diminish) decision makers' concerns about how prosocial they are perceived to be, while their concerns of how influenceable they are perceived to be do not change.

5.3. Narrative usage

To study the use of narratives, we exploit the fact that every subject in the main experiment was asked to provide an explanation for her choice at the very end of the experiment.[33] 122 raters classified these narratives into different categories (see Table C.4 for a full list). We, thus, obtain 609 narratives that are linked to at least one category (11% are not classified by our procedure). Overall, the most prominent category is fairness, i.e., narratives that explicitly make a statement about fairness or egalitarian principles. The second and third most prominent types of narratives are ones explicitly mentioning selfishness as an explanation for one's behavior and narratives referring to the fact that both participants are equally deserving of the money. Interestingly, very few subjects explicitly mention the narratives used for the treatment manipulations in their own explanation (2% in the Positive and 4% in the Negative condition).

In the online experiment, the narratives we used in the Positive condition are both classified in the fairness category, while the ones used in the Negative condition both use anonymity as an explanation and one additionally explicitly refers to the need for money. We, hence, look at whether the occurrence of these categories is influenced by the treatment conditions. Holding a certain action constant (giving 5 or giving 0), there is no difference in the types of narratives used between the Positive condition and the Baseline (Fisher's exact tests, all $p > 0.1$). The Negative condition decreases the likelihood of using narratives relating to fairness to justify giving 5 (Negative vs. Baseline, 80% vs 90%, Fisher's exact-test, $p = 0.068$), while there is no difference for other categories. Considering subjects who give 0, negative narratives increase the likelihood of them referring to anonymity to justify their behavior (Negative vs. Baseline, 25% vs. 11%, Fisher's exact-test, $p = 0.088$). Again there is no difference for the other categories. Last, it is interesting to note that narratives are quite homogeneous across selfish and prosocial types. In other words, the narratives that subjects make use of to justify a certain choice (e.g., giving 5 or giving 0) do not vary systematically between types.

Result 9. The categories of narratives used to explain a given choice are quite homogeneous across types and treatments. Narratives refer less often to fairness to justify giving 5 in the Negative condition compared to the Baseline. Narratives mention anonymity more often to justify giving 0 in the Negative condition compared to the Baseline.

6. Discussion and conclusion

Our results provide insights into how narratives in favor of prosocial or selfish actions influence the behavior of different social types. In our main experiment, subjects see either positive or negative narratives upon taking a distributional choice in a dictator game. We compare our two treatment conditions with a Baseline in which no narratives are provided. Empirical beliefs about the distribution of choices are fixed across all experimental conditions. We run two versions of our main experiment, one in the lab and a robustness check online. We work out two hypotheses from a theoretical framework that models how narratives influence behavior via the perception of the appropriateness of an action or the deservingness of a

[32] Note that subjects in all experimental conditions were shown the distribution of choices of other subjects, making the notion of influenceability also viable in the Baseline condition.

[33] We only use narratives from the online experiment for this analysis, as the ones from the lab stem from a different experimental setting and a different subject pool (German university students). Hence, they would not necessarily make sense for the raters used in this experiment (US Prolific subjects) and nuances may get lost in translation.

recipient for different social types. We complement this evidence with four additional experiments, which help us to shed some more light on the mechanisms through which narratives work and provide a more encompassing picture of these phenomena.

In both the lab and online experiment, we find that subjects in the Positive condition give more than subjects in the Baseline condition. This increase is predominantly driven by selfish types and leads subjects to give 0 less often than in the Baseline. Coherently with these findings, our additional experiments show that positive narratives increase the appropriateness of giving 5, but — even more strongly — decrease the appropriateness of giving 0. In addition, they also affect the way giving 0 is perceived by making a decision maker look more selfish than someone who does the same when there are no narratives. Positive narratives are also effective in increasing a decision maker's own concerns about how prosocial their actions are perceived to be. Taken together these findings imply that positive narratives are effective at changing the perception of different choices and lead to more prosocial behavior by making selfish choices less attractive.

Negative narratives are in some way also effective in shifting these perceptions. However, they do not influence behavior in the expected direction. We find no aggregate effect of the Negative condition neither in the lab nor in the online experiment. Only in the lab do we find that selfish types give more and very prosocial types give less compared to the Baseline. These effects are surprising at first glance because the presence of negative narratives increases the appropriateness of giving 0 and — as above — cause an even stronger decrease in the appropriateness of giving 5. This is also reflected in the fact that subjects who give 5 in the Negative treatment are less likely to use fairness arguments in their own narratives. Moreover, someone giving 0 is perceived as more prosocial compared to the case without narratives. Finally, decision makers' own concerns about how prosocial their actions are perceived to be become weaker in this treatment condition.

This implies that the mechanisms described by Bénabou et al. (2020) are confirmed by our data. The presence of narratives effectively influences the perceived appropriateness of giving. At the same time, negative narratives allow selfish subjects to pool with prosocial subjects. However, the question that remains is why only positive narratives are effective in changing behavior.

One element that can explain this, is the finding that someone who follows the narrative (gives 5 in the Positive condition or gives 0 in the Negative condition) is perceived as more influenceable. In contrast to other image concerns, the relative importance of not seeming influenceable decreases in the Positive and increases in the Negative condition. Hence, subjects in the Negative condition could shy away from acting selfishly to avoid being perceived as influenceable. In line with this explanation, very few subjects explicitly mention the narratives used for the treatment manipulations in their own explanations (2% in the Positive and 4% in the Negative condition).

These findings illustrate that, while negative narratives may help decision makers to hide behind them, they also come at a cost, since some undesirable traits may be attached to someone acting in line with them. This may lead decision makers not to buy into the narrative, and, under certain circumstances, it may even lead them to give more, as we find in our lab experiment. However, it is important to note, that a causal analysis of the relationship between influenceability and behavior would be needed to reach a definitive conclusion. Unfortunately, little is known about how perceptions of influenceability affect economic decisions, and a thorough investigation of this mechanism falls beyond the scope of this paper. However, our results indicate that positive and negative narratives are not just the flip sides of the same coin.

Our findings also relate to a literature on the 'focusing effect', a term coined by Krupka and Weber (2009), who find that descriptive information enhances prosocial behavior, even in cases where one does not observe a lot of norm-compliant behavior.[34] Similarly, Gino et al. (2009) find that increasing the saliency of an opportunity to cheat decreases unethical behavior. This resonates also with a study by Xiao (2017) who shows that the pressure to justify leads to more norm-compliant behavior in prosocial choices. In this sense, the moral salience induced by narratives might lead 'reluctant sharers' to give (Lazear et al., 2012). This suggests that narratives — regardless whether positive or negative — may enhance the moral saliency of the decision.

Our work advances the understanding of the determinants of prosocial behavior by providing insights into how narratives — which permeate people's life — work. Our findings suggest that narratives change how people perceive a certain decision but do not necessarily change behavior. In this sense, existing models need to be enriched to account for multiple, potentially counteracting motives that may influence behavior. Our results also have relevant implications for institutions and organizations that can use narratives to promote prosocial behavior, especially amongst the people who would be less inclined to act so ex ante. According to our findings even the presence of some negative narratives would not necessarily harm prosocial behavior.

An important contribution of our work is that it brings to light a subtle difference between positive and negative narratives. More research is needed to fully understand this difference. A possible direction could be that of exploring negative narratives as excuses, which may work on the surface, but may in some sense backfire by making the moral content of a decision more salient. In the same vein, a peculiarity of the dictator game is that sharing the money equally represents a clear norm. Future research could investigate the relationship between narratives and the strength of a norm or the presence of multiple norms. Other questions are how enduring the effect of a certain narrative is, and whether there might be spillovers in other contexts. We hope our work can contribute to inspire such endeavors.

[34] A similar effect can be found in Berg et al. (1995), who study a trust game in which showing trustees selfish behavior from previous rounds increases their back-transfers.

Declaration of competing interest

There are no conflicts of interest.

Availability of Data and Material

The data and materials are available from the authors.

Code Availability

The code of the program is available from the authors.

Funding

Funding was granted by the Max Planck Society.

Appendix A. Theoretical framework

This section complements the "Behavioral Predictions" in the main text (Section 3.3) by providing formal definitions and derivations of the hypotheses. A decision maker chooses how much money to give to a recipient. A key component of this model is the belief about the externality of giving (Bénabou et al., 2020). We, first, describe the basic utility function of a decision maker; we, then, explain which role the externality plays; and, then, discuss how narratives enter the model.

The utility function of a decision maker (DM) takes the following form:

$$U_i(g,e) = v(g,e) - c(g), \tag{1}$$

where g is the amount she decides to give, and e is the expected externality of giving, which we define below; $v(g,e)$ captures the overall valuation of giving, and $c(g)$ the costs of giving.[35] We set $e \in (0,1)$ and assume $c(g)$ to be linear increasing in g. While $v(g,e)$ can take many functional forms, we assume that v is increasing and concave in g ($\frac{\partial v(g,e)}{\partial g} > 0$, $\frac{\partial^2 v(g,e)}{\partial g^2} < 0$). This assumption ensures an internal solution with an optimal amount of giving $g^*(e)$.

The externality. E is a binary measure of the presence of a positive externality, i.e., whether the recipient is deserving or it is appropriate to give in the situation at hand (see discussion in Section 3.3). If $E=1$, there is a positive externality, while if $E=0$, there is no such externality. A DM in our model does not know the value of E with certainty. Rather, she holds a prior belief (what we call perception above) about E with $e = P(E=1)$. We assume that the marginal utility of giving is increasing in the expected externality e ($\frac{\partial v(g,e)/\partial g}{\partial e} > 0$). Following this assumption, a higher e leads to higher amounts of giving. Note that $v(g,e)$ can take on many different forms. In a setting like the standard dictator game the strong focal point at the equal split could be understood as a norm. Correspondingly, in an alternative modeling approach, setting $v(g,e) = -\gamma(e)(\frac{1}{2} - g)^2$ in a dictator game with a pie size of 1, $\gamma(e)$ would capture the appropriateness to follow the norm, i.e., to split the pie equally (assuming $\frac{\partial \gamma}{\partial e} > 0$). Independently of the specific choice of v, our predictions hold.

Narratives. We model narratives as signals about E updating the prior belief of a DM, as in Bénabou et al. (2020). A positive narrative signals that $E=1$, i.e., it is an argument or justification for there being a positive externality. A negative narrative, conversely, signals that $E=0$. For simplicity, we take DMs to be standard Bayesian updaters. Other forms of updating are of course conceivable, but would introduce further degrees of freedom in the model. Moreover, as long as an alternative updating model leads to updating in the same direction for all priors and leads to different posteriors for different priors, the main intuitions of the model will hold. We assume narratives to be at least somewhat believable or convincing, which here means that the signal is correct more often than not. Hence, a DM will update in the direction of the signal.[36]

As an example, let us assume a signal structure as in Fig. A.1. If there is no externality $E=0$, with probability $1 \geq s > \frac{1}{2}$ the correct signal, i.e. the negative narrative, is sent, and with $1-s$ the signal is wrong, i.e. the narrative is positive. The situation is reversed with a high externality ($E=1$).

[35] Note that all factors influencing the utility of giving are captured by the first term. For the sake of simplicity, we do not consider how image concerns would alter the resulting trade-off.

[36] Note that Bénabou et al. (2020) formally define positive and negative narratives directly by their influence on beliefs. The signaling structure we use is based on an older version of their paper and leads to the same directional effect of narratives on actions.

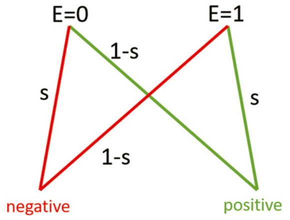


Fig. A.1. Exemplary signal structure.

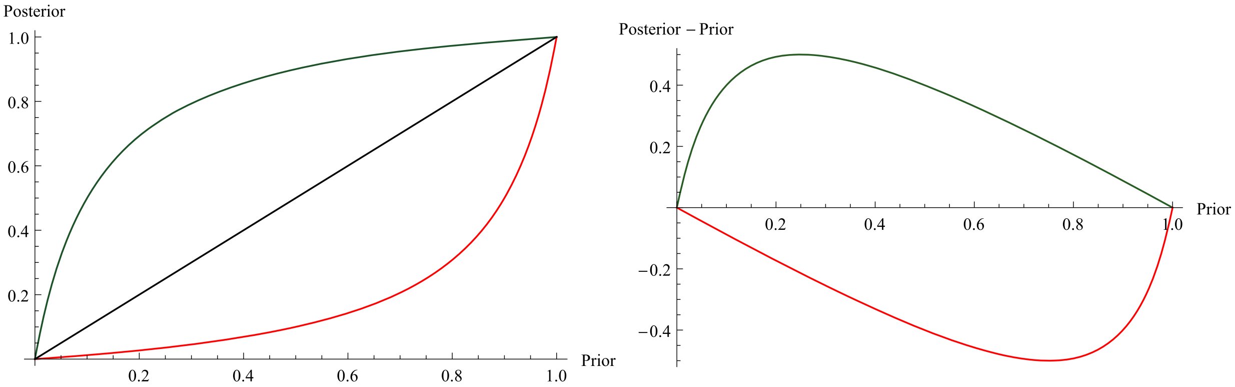


Fig. A.2. Posterior for given signal. Note: The left figure shows posterior beliefs as a function of prior beliefs and the right figure shows the corresponding difference between posterior and prior beliefs, both after receiving a positive (green, upper line) or negative signal (red, lower line), dependent on the prior belief. For these examples, we set $s = 0.9$. The black line on the left is the 45-degree line representing the case with no signal or no updating.

The posterior given a positive or negative signal is calculated as follows (with e being the prior probability of $E = 1$). Fig. A.2 provides a graphical representation.

$$P_{post}(E=1|Positive) = \frac{P(Positive|E=1)P_{prior}(E=1)}{P(Positive)} = \frac{se}{se+(1-s)(1-e)}$$

$$P_{post}(E=1|Negative) = \frac{P(Negative|E=1)P_{prior}(E=1)}{P(Negative)} = \frac{(1-s)e}{(1-s)e+s(1-e)}$$

Given this signal structure, negative narratives lead to a downward shift in beliefs and positive narratives to an upward shift. That is, independent of the prior belief, the posterior belief is decreasing when receiving a negative narrative and increasing when receiving a positive narrative for the full range of beliefs. Since, as stated above, higher beliefs about e translate into higher amounts of giving, our first hypothesis follows directly.

Hypothesis 1. Positive narratives increase giving, while negative narratives decrease giving.

Heterogeneity. We introduce heterogeneity by allowing diverging beliefs about E.[37] In fact, DMs in our model differ solely in their beliefs, which we bound to $e \in (0, 1)$. That is, all DMs in our model would act in the same way, i.e., choose to give the same amount, if they held the same belief. Modeling heterogeneity solely through beliefs offers us a concise way to introduce narratives as signals. We call DMs with low beliefs ‘selfish’ types and those with high beliefs ‘prosocial’ types.

While in our framework the direction of the effect of narratives is independent from prior beliefs, our setup predicts a different strength of the effect for different priors. In particular, extreme types (those with priors $\hat{e}$ close to 0 or close to 1) will not update strongly when receiving a signal close to their prior belief, whereas they will update strongly when receiving a contradicting signal (Fig. A.2).

Hypothesis 2. Positive narratives should have a stronger positive effect on more selfish types, while negative narratives should have a stronger negative effect on more prosocial types.

[37] Bénabou et al. (2020) hint at heterogeneity in priors, but consider common priors throughout the paper with heterogeneity between subjects stemming solely from different valuations of the externality.

Appendix B. Additional analysis main experiment

B.1. Main results

This section contains complementary information to our main analysis in Section 4. In Table B.1 we report the results of our main regression. The visualization of these results is provided in Fig. 2 in the main text. In Table B.2, we report the results of Probit regressions with giving 5 and giving 0 as dependent variables in the lab and the online experiment. These results are visualized in Fig. 3.

Table B.1
Tobit regressions.

	(1) Lab	(2) Lab	(3) Online	(4) Online
POSITIVE	0.752**	2.852***	0.666***	0.714
	(0.360)	(0.888)	(0.256)	(0.667)
NEGATIVE	0.125	2.698***	-0.0553	-0.216
	(0.360)	(0.894)	(0.253)	(0.598)
type	0.133***	0.189***	0.0979***	0.0963***
	(0.0116)	(0.0217)	(0.00824)	(0.0134)
POSITIVE x type		-0.0732**		-0.00140
		(0.0283)		(0.0207)
NEGATIVE x type		-0.0900***		0.00566
		(0.0285)		(0.0191)
Constant	-1.382***	-3.015***	0.169	0.216
	(0.428)	(0.696)	(0.297)	(0.420)
Observations	280	280	682	682
Pseudo R^2	0.108	0.118	0.051	0.051

Standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.
Note: Tobit regression with lower censoring at 0 (84 censored observations in lab and 151 in online experiment). The type measure corresponds to the SVO angle, Positive and Negative conditions are included as dummies. We also include interaction terms between treatment conditions and the SVO angle in column (2) and (4).

Table B.2
Probit regressions, give 0 and give 5.

	give 0		give 5	
	(1)	(2)	(3)	(4)
	Lab	Online	Lab	Online
POSITIVE	-0.975**	-0.413	1.578***	-0.0718
	(0.463)	(0.334)	(0.547)	(0.352)
NEGATIVE	-1.230***	0.0467	1.043*	-0.353
	(0.472)	(0.280)	(0.598)	(0.329)
type	-0.0825***	-0.0335***	0.0820***	0.0361***
	(0.0143)	(0.00693)	(0.0142)	(0.00749)
POSITIVE x type	0.0204	-0.00288	-0.0405**	0.00822
	(0.0193)	(0.0119)	(0.0179)	(0.0112)
NEGATIVE x type	0.0491***	-0.00746	-0.0346*	0.00465
	(0.0180)	(0.00985)	(0.0189)	(0.0103)
Constant	1.617***	0.299	-2.705***	-1.030***
	(0.352)	(0.198)	(0.443)	(0.235)
Observations	280	682	280	682
Pseudo R^2	0.275	0.139	0.205	0.122

Standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.
Note: Probit regression. The dependent variable is giving 0 in the first two columns and giving 0 in the third and fourth column. The type measure corresponds to the SVO angle, Positive and Negative conditions are included as dummies. We also include interaction terms between treatment conditions and types.

B.2. Robustness checks

In Table B.3 we conduct multiple robustness checks with our lab data. In the first two columns we control for the additional psychological measures and sessions.[38] In column 3, we impose both lower and upper censoring. Column 4

[38] We do not include demographics as controls since they were only recorded for 22 subjects in the Baseline treatment and the comparison would thus be underpowered.

Table B.3
Robustness checks - lab.

	(1)	(2)	(3)	(4)	(5)
	Tobit	Tobit	Tobit, upper and	Tobit	OLS
	controls	sessions	lower censoring	quadratic	
POSITIVE	2.419***	1.884*	5.799***	6.856	1.468***
	(0.855)	(1,062)	(2.166)	(4.548)	(0.555)
NEGATIVE	2.635***	2.709**	5.494**	11.96**	1.313***
	(0.868)	(1.086)	(2.162)	(3.907)	(0.560)
Type	0.165***	0.163***	0.405***	38.78***	0.123***
	(0.0211)	(0.0211)	(0.0613)	(12.77)	(0.0133)
POSITIVE x type	-0.0580**	-0.0532**	-0.142**	-15.32	-0.0365**
	(0.0270)	(0.0269)	(0.0717)	(16.54)	(0.0187)
NEGATIVE x type	-0.0905***	-0.0918***	-0.194***	-36.41**	-0.0487***
	(0.0275)	(0.0275)	(0.0717)	(14.22)	(0.0188)
Type^2				-21.78**	
				(10.74)	
POSITIVE x type^2				8.518	
				(13.99)	
NEGATIVE x type^2				26.02**	
				(12.16)	
Constant	-3.685**	-3.5625*	-7.972***	-12.83***	-0.509
	(1.867)	(1.9015)	(1.815)	(3.558)	(0.397)
Controls	yes	yes	no	no	no
Session	no	yes	no	no	no
Observations	280	280	280	280	280
Pseudo R^2	0.144	0.1512	0.140	0.124	0.3647

Standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.
Note: Tobit and OLS regressions. The type measure corresponds to the SVO angle, Positive and Negative conditions are included as dummies. We also include interaction terms between treatment conditions and types. Controls include Context Dependence, Context Independence, Moral Identity Scale, Moral Disengagement, and the 11-item, Big-5 questionnaire. Session dummies are used as controls. In the Tobit model in column (3) 84 observations are censored at giving 0 and 120 observations censored at giving of 5.

introduces a quadratic term for the type measure and for its interactions with the treatment conditions. Fig. B.1 shows the marginal effects for columns 2-4. The pattern described in Section 4 remains substantively unchanged for all these alternative specifications. In column 5, we run a standard OLS regression. Also in this case, results are comparable to those of our main regressions.

In Table B.4, we conduct the same robustness checks with our online data, except for the one in which we only include sessions as controls. Fig. B.2 shows the marginal effects for columns 1-3. Again the results are not substantially different than in our main analysis.

B.3. Analysis of additional psychological measures (lab only)

In Table B.5, we run the same analysis as in Table B.1 using the additional psychological measures collected in the type elicitation experiment. Both Moral Identity and Moral Disengagement have a strong and highly significant relationship with giving in the expected direction, i.e., positive and negative, respectively. However, they do not contribute significantly to the explanation of our treatment effects. Meaning that the Negative and Positive condition do not affect subjects scoring differently on these scale in a different way. This gives us further assurance in using the incentivized SVO measure for the main analysis. As to the complementary measures of Context Dependence and Independence, they do not significantly mediate our treatment effects. Meaning that the treatment conditions do not affect subjects who are more or less dependent from the context in making their decisions, as measured by these scales, differently.

B.4. Feelings (lab only)

In Table B.6, we regress the measures of feelings we collected after subjects' choice in the dictator game. In all columns, we regress a specific measure on dummies for treatment conditions, the amount a subject gave, her SVO angle and an interaction term between the latter and the treatment conditions. The first two columns refer to general feelings of happiness and contentment (how happy/contented do you feel at the moment?), which are rather stable. The last four columns refer to feelings regarding a subject's choice in the dictator game. Guilt and shame decrease in the amount a subject gives. However, the presence of negative or positive narratives in our treatment conditions does not substantially alter this relationship. Nevertheless, we cannot rule out that the absence of treatment effects is caused by the anticipation of these feelings. The presence of narratives could lead subjects to anticipate guilt or shame and to adapt their giving to avoid them, which could result in similar stated feelings across treatments.

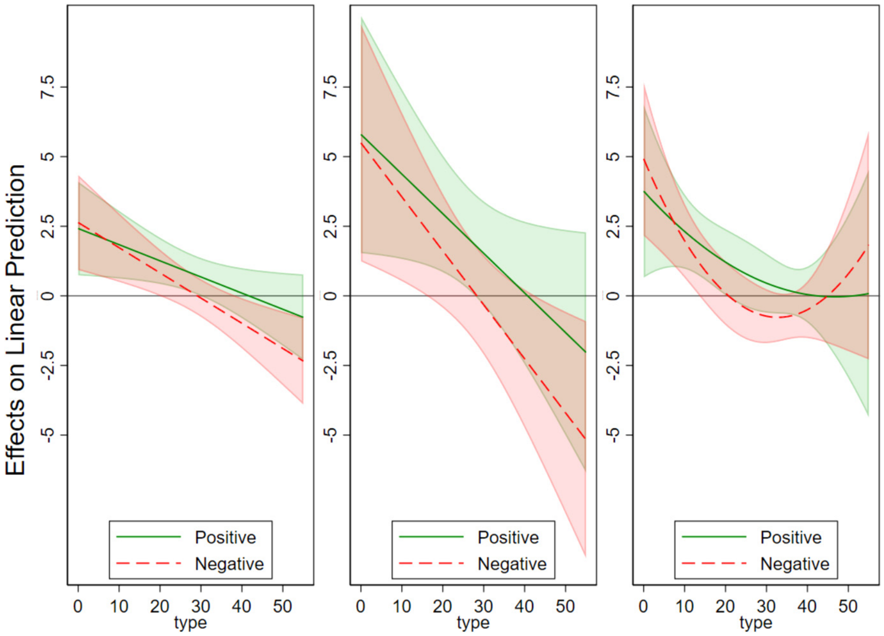


Fig. B.1. Marginal effects, Tobit - lab. Note: Tobit with lower censoring at 0 and controls on the left, Tobit with upper and lower censoring (5 and 0) in the middle. Tobit with quadratic interaction term on the right. For the ease of visualization, types below 7° (3 subjects) are not displayed. Outer lines show 95%-confidence intervals.

Table B.4
Robustness checks - online.

	(1) Tobit controls	(2) Tobit, upper and lower censoring	(3) Tobit quadratic	(4) OLS
POSITIVE	0.665	0.685	-1.057	0.281
	(0.662)	(1.784)	(0.974)	(0.509)
NEGATIVE	-0.293	-0.587	-0.549	-0.202
	(0.600)	(1.575)	(0.746)	(0.446)
Type	0.0910***	0.252***	0.0635	0.0719***
	(0.0134)	(0.0397)	(0.0449)	(0.0102)
POSITIVE x type	0.000642	0.0495	0.190**	0.00657
	(0.0206)	(0.0590)	(0.0810)	(0.0160)
NEGATIVE x type	0.00909	0.00314	0.0546	0.00345
	(0.0191)	(0.0527)	(0.0666)	(0.0146)
Type^2			0.000724	
			(0.000953)	
POSITIVE× Type^2			-0.00383**	
			(0.00159)	
NEGATIVE × Type^2			-0.00107	
			(0.00139)	
Constant	7.505	-1.956*	0.432	1.325***
	(10.81)	(1.130)	(0.499)	(0.312)
Controls	yes	no	no	no
Observations	674	682	682	682
Pseudo R^2	0.054	0.084	0.053	0.1865

Standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.
Note: Tobit and OLS regressions. The type measure corresponds to the SVO angle, POSITIVE and NEGATIVE conditions are included as dummies. We also include interaction terms between treatment conditions and types. Controls include charitable giving, social comparison, age, gender, Prolific score, and number of approvals on Prolific. In the Tobit model in column (2) 151 observations are censored at giving 0 and 373 observations censored at giving of 5.

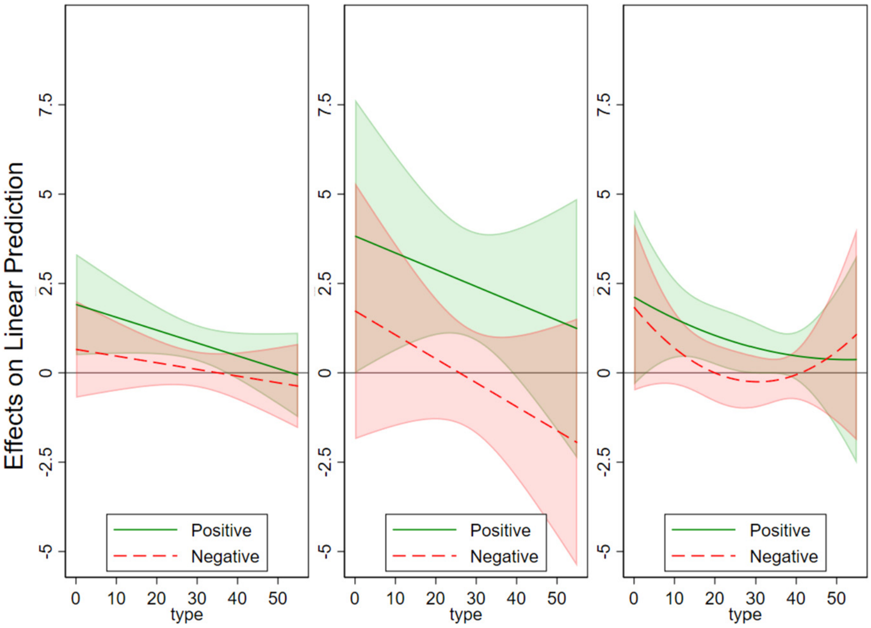


Fig. B.2. Marginal effects, Tobit - online. Note: Tobit with lower censoring at 0 and controls on the left, Tobit with upper and lower censoring (5 and 0) in the middle. Tobit with quadratic interaction term on the right. For the ease of visualization, types below 7° (20 subjects) are not displayed. Outer lines show 95%-confidence intervals.

Table B.5
Tobit regressions, alternative measures.

	Moral identity	Moral disengagement	Context dependence	Context independence
POSITIVE	1.500	1.705	1.485	1.391
	(2.340)	(1.933)	(1.352)	(2.185)
NEGATIVE	0.308	0.823	-0.0235	0.495
	(2.399)	(2.053)	(1.402)	(2.171)
measure	1.303***	-1.222**	-0.0443	0.116
	(0.401)	(0.489)	(0.243)	(0.412)
POSITIVE × measure	-0.270	-0.274	-0.211	-0.188
	(0.567)	(0.676)	(0.344)	(0.583)
NEGATIVE × measure	-0.133	-0.248	0.0251	-0.117
	(0.581)	(0.735)	(0.352)	(0.587)
Constant	-2.913*	5.506***	2.329**	1.738
	(1.613)	(1.349)	(0.952)	(1.538)
Observations	280	280	280	280
Pseudo R^2	0.024	0.023	0.004	0.003

Standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.
Note: Tobit regressions with lower censoring at 0. The type measure corresponds to the stated measure, POSITIVE and NEGATIVE conditions are included as dummies. We also include interaction terms between treatment conditions and the SVO angle.

B.5. Social comparison (online only)

We test whether our social comparison measure explains giving. This is done to test the hypothesis that narratives serve as a benchmark against which subjects compare themselves. Column (1) - (3) in Table B.7 consider a Tobit regression with lower censoring at 0. Columns (4) - (7) use giving 5 and giving 0 as dependent variable. We analyze the behavior of prosocial and selfish types separately. The results show that social comparison has no effect on giving or on the probability of following the narrative. The same holds true when controlling for the SVO angle. Note that we also include social comparison as a control variable in the robustness check in Table B.4. There, it also has no effect on giving.

Table B.6
OLS regressions, feelings.

	Happiness	Content	Guilt	Contentment	Shame	Excited
Constant	4.137***	3.854***	2.440***	4.169***	2.089***	2.598***
	(0.319)	(0.331)	(0.264)	(0.261)	(0.229)	(0.326)
Positive	0.694	0.756	0.455	0.318	0.240	0.553
	(0.451)	(0.468)	(0.373)	(0.369)	(0.323)	(0.461)
Negative	0.651	1.034*	-0.127	0.454	0.246	-0.027
	(0.454)	(0.470)	(0.376)	(0.371)	(0.325)	(0.464)
Type	0.013	0.017	0.012	0.018	0.005	0.001
	(0.012)	(0.013)	(0.010)	(0.010)	(0.009)	(0.013)
Give	-0.003	0.040	-0.309***	0.001	-0.213***	0.032
	(0.048)	(0.050)	(0.040)	(0.040)	(0.035)	(0.050)
Positive × Type	-0.012	-0.019	-0.014	-0.012	-0.008	-0.018
	(0.015)	(0.016)	(0.012)	(0.012)	(0.011)	(0.015)
Negative × Type	-0.017	-0.023	0.008	-0.017	-0.002	0.000
	(0.015)	(0.016)	(0.013)	(0.012)	(0.011)	(0.016)
Adj. R^2	-0.004	0.009	0.210	-0.005	0.162	-0.012
Num. obs.	280	280	280	280	280	280

Standard errors in parentheses.
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.
Note: OLS regressions. The dependent variables are the stated feelings. The first two columns refer to general feelings, the last four columns refer to feelings specific to the choice. The type measure corresponds to the SVO angle, giving is the amount given, Positive and Negative conditions are included as dummies. We also include interaction terms between treatment conditions and the SVO angle.

Table B.7
Tobit regressions, social comparison analysis.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	give	give	give	give 0	give 0	give 5	give 5
		selfish	prosocial	selfish	prosocial	selfish	prosocial
positive	-0.818	0.595	-0.681	-0.952	0.756	-0.728	0.806
	(1.395)	(3.601)	(1.344)	(1.302)	(0.895)	(1.448)	(0.721)
negative	-0.501	-0.230	0.544	-0.551	0.283	1.298	-0.338
	(1.569)	(3.612)	(1.603)	(1.056)	(0.803)	(1.193)	(0.715)
SC	-0.537	-0.449	-0.214	0.129	0.0992	0.175	0.0841
	(0.445)	(0.991)	(0.432)	(0.293)	(0.214)	(0.331)	(0.212)
positive x SC	0.734	0.291	0.540	0.166	-0.543	0.356	-0.240
	(0.561)	(1.441)	(0.538)	(0.516)	(0.380)	(0.569)	(0.294)
negative x SC	0.191	0.183	-0.310	0.187	-0.167	-0.580	0.0326
	(0.635)	(1.495)	(0.640)	(0.421)	(0.327)	(0.479)	(0.291)
Constant	4.160***	1.409	4.295***	-0.256	-1.166**	-1.239	0.0807
	(1.106)	(2.558)	(1.073)	(0.758)	(0.533)	(0.865)	(0.523)
Observations	682	190	492	190	492	190	492
Pseudo R^2	0.006	0.005	0.006	0.029	0.029	0.019	0.021

Standard errors in parentheses.
* $p < .10$, ** $p < .05$, *** $p < .01$.
Note: Tobit with lower censoring at 0. SC refers to our social comparison measure. The type measure corresponds to the SVO angle, Positive and Negative conditions are included as dummies. We also include interaction terms between treatment conditions and the social comparison measure.

Appendix C. Additional experiments

Here we provide more details about the experiments described in Section 5. We also provide additional information regarding the results of these experiments. The procedure adopted for the recruitment of subjects is the same as the one used for the online experiment (see Section 3.2). Note that all participants in the additional experiments were selected as recipients of the main experiment. However, they learned about this only at the very end of their experiment and after they made all choices.

C.1. Appropriateness

In this experiment, subjects had to evaluate how socially appropriate it is to give 0 or 5 in our three experimental conditions (Baseline, Positive and Negative). They were incentivized to guess the modal answer of all other participants in the experiment (see Krupka and Weber, 2013). These incentives create a coordination game in which the social norm can act as a focal point. Answers could range from not appropriate at all to very appropriate on a 6-point likert scale. Subjects first went through a description of the incentive mechanism and then a description of the dictator game implemented in our main experiment. The experiment was run in a within-subjects design, i.e., all subject went through the three experimental conditions. They saw the same screen in which subjects took their decision in the main experiment (see Fig. D.2). In the end, one situation was selected randomly to determine payments. Subjects received £0.50 for participation and could earn an additional £1 in case their guess corresponded to the modal answer. We obtained data from 179 subjects (4% failed the attention check).

C.2. Image concerns

To investigate image concerns, we run two separate experiments: one on how decision makers giving either 0 or 5 in the three experimental conditions are perceived and one on how important these perceptions are to the decision makers themselves in the three experimental conditions.

In the first experiment, subjects first saw a description of the dictator game implemented in our main experiment. Then, they had to guess the SVO-angle of a decision maker who gave 0 or 5 in the three different experimental conditions. Their guess could range from 1 to 10 and each number corresponding to the deciles of the type distribution observed in the experiment (e.g., stating 4 means that the subject was in between the 30th and 40th percentile of the distribution). This measure was incentivized. At the end of the experiment one of the six guesses was randomly selected and subjects were paid according to how close their guess was to the actual SVO-angle of a randomly drawn subject who gave either 0 or 5 in that treatment condition. We used the binarized scoring rule (Hossain and Okui, 2013) to ensure incentive compatibility. Subjects could earn either £1 or £0 depending on their guess. As suggested by Danz et al. (2020), we did not explain the payments scheme in detail to subjects but told them that it was in their best interest to provide their best guess. We also elicited an unincentivized measure of confidence about subjects' guess. Additionally, we asked subjects to evaluate how trustworthy, likable and influenceable a decision maker in each situation was (on a 7-point Likert Scale). These measures were not incentivized. This experiment was conducted in a within-subjects design and the order of appearance of the three different experimental conditions was randomized at the individual level. Subjects received £0.50 for participation and earned an additional £1 if their guess exactly right. 101 subjects completed the experiment (9% failed the attention check).

In the second experiment, subjects also saw a description of the Dictator Game, but this time they were asked to answer a series of questions as if they were the decision makers and would decide how much to give. We asked subjects to state how much they would care about being seen as prosocial, trustworthy, likable and not influenceable in the three treatment conditions (on a 7-point Likert Scale). We also asked them how much they would care about the opinion of others in general in each situation (also, on a 7-point Likert Scale). Subjects also participated in the type elicitation experiment, just like subjects in the main experiment. Also in this case, the order of appearance of the three treatment conditions was randomized at the individual level. They received £0.75 for participation. 92 subjects completed the experiment (6% of potential participants failed the attention check).

C.2.1. Additional results

Table C.1 and Table C.2 show the scores assigned by subjects for how prosocial, trustworthy, likeable and influenceable a decision maker who gave 0 or 5 was according to them. Table C.3 shows the scores assigned to the importance of these traits by decision makers themselves.

Table C.1
How are decision makers perceived by others - give 0.

give 0	Baseline	Positive	Negative
prosocial	2.31	2.11*	2.94***
trustworthy	2.50	2.47	2.82***
likeable	2.47	2.32**	2.69*
influenceable	3.16	2.92	3.57***

Note: Wilcoxon signrank-test in comparison to Baseline.
(***: $p < 0.01$, **: $p < 0.05$ *: $p < 0.1$).

Table C.2
How are decision makers perceived by others - give 5.

give 5	Baseline	Positive	Negative
prosocial	7.54	8.00***	7.32
trustworthy	5.61	5.80**	5.59
likeable	5.71	6.38***	5.61
influenceable	4.41	4.72**	4.19

Note: Wilcoxon signrank-test in comparison to Baseline.
(***: $p < 0.01$, **: $p < 0.05$ *: $p < 0.1$).

Table C.3
Importance of image concerns for decision makers.

	Baseline	Positive	Negative
prosocial	3.92	4.36***	3.59*
trustworthy	5.10	5.54***	4.77*
likeable	4.55	4.72	4.01***
not influenceable	3.37	3.22	3.39
social image	3.80	4.16***	3.29***

Note: Wilcoxon signrank-test in comparison to Baseline.
(***: $p < 0.01$, **: $p < 0.05$ *: $p < 0.1$).

Table C.4
Categories.

	Baseline	Positive	Negative
Positive narratives			
Deservingness	13	18	13
Golden rule	7	7	8
Fairness	47	62	39
Empathy guilt	4	2	6
Kindness	9	8	13
Negative narratives			
Anonymity	4	1	9
Need money	11	9	10
Unknown recipient	10	6	15
Selfishness	19	13	15
Other narratives			
Descriptive norm	6	6	6
Graph	13	4	5
Narratives	0	2	4
Other	0	0	1

Note: Narrative frequency (in percent). A narrative can be linked to multiple categories.

Table C.5
Description of categories.

Category	Description
Positive narratives	
Deservingness or equal effort:	The Recipient is similarly deserving or has put similar effort compared to the Decider.
Karma or golden rule:	The Decider wants to treat others as he or she would like to be treated by them or wants to increase his or her karma.
Equality or fairness:	The Decider wants to behave according to principles of equality or fairness.
Empathy or guilt:	The Recipient expects something or would be disappointed from receiving nothing.
Kindness or generosity:	The Decider wants to be kind or generous.
Negative narratives	
Anonymity:	The Recipient does not know the Decider's identity.
Need for money:	The Decider needs the money for him or herself.
Unknown identity of other participant:	The Decider does not know who the Recipient is.
Selfishness or greed:	The Decider declares to be selfish or greedy.
Other narratives	
Descriptive Norm	Belief or expectation that others will/would behave in the same way in the situation.
Graph	Reference to how other participants behave according to the graph.
Narrative	Reference to the explanations of other participants.
Other:	Please describe in your own words a suitable category.

C.3. Narrative usage

In this experiment, 219 subjects categorized the narratives from our main experiment. Subjects first saw a description of the Dictator Game. Then, they were presented with the available categories (see Table C.5). We discarded incomprehensible or missing narratives and ended up with 683 narratives. Each subject categorized 32 randomly chosen narratives, some of which were attention checks to make sure that subjects responses were reliable (about 2% of participants failed the attention checks). A narrative is classified in a certain category if more than half of the raters put it in that category. Hence, a single narrative can be classified in multiple categories. Of the 683 narratives, 74 (10.83%) were not clearly classified with our procedure and are thus excluded from the analysis leaving us with 609 narratives in total. Table C.4 shows the usages of narratives in the different conditions.

Appendix D. Additional materials

This Appendix contains additional information about the lab and online experiment.

D.1. Lab experiment

D.1.1. Narrative selection

Table D.1 shows positive and negative narratives (translated from German) along with their average convincingness rating. Numbers 1-4 were selected for the Positive condition and 5-8 for the Negative condition. 3 raters rated a total of 73. Narratives were selected from the first 3 sessions of the Baseline condition, since the 4th session was run later to balance the number of participants in all conditions. A complete list of narratives is available from the authors upon request.

Table D.1
Narrative selection lab.

Number	Positive Narratives	Convincingness
1	Both came here to participate in the experiment and spent the same amount of time here. Both should get the same payment.	6
2	An equal distribution of the money is only logical: Assuming everyone agrees on that, everyone will go home with 10 €. Everything else would be a mixture of greed and speculation.	6
3	Fair choice. Everyone gets exactly the same amount of money. Since it is unknown who Person B is and whether her life circumstances would justify another distribution, this is the only just decision.	6
4	I think that both participants should get the same amount of money. If it is unknown in advance whether you are A or B it is just smart to give 5 € to both.	6.3
	Negative Narratives	
5	Since the experiment is anonymous, I expect that everyone is looking for her own advantage. I don't know any of the other players and since the decision happens randomly anyway, I do not care about giving someone else money.	6
6	This way I get the highest payoff in case I am participant A. In case I am participant B, I have no influence on my payoff because of the assignment to role B.	5.6
7	Because I would like to have the money and saw in the statistic that others also decided this way. This made me have less scruples for allocating all the money to myself.	5.3
8	I allocated 10 € to myself, since this way I get the most money on average. As it is unclear how much I would get as participant B, I wanted to achieve the maximum profit in case I am participant A.	5.3

D.1.2. Additional psychological measures

We elicited various psychological measures at the end of the experiment. We include the 11-item, Big5 questionnaire (Rammstedt and John, 2007), the Context Dependence and Independence questionnaire (Gollwitzer et al., 2006), a reduced form of the Moral Disengagement questionnaire (Bandura et al., 1996), and a modified version of the Moral Identity Scale (Aquino and Reed, 2002).

Big 5 Questionnaire. This questionnaire is taken from Rammstedt and John (2007).

Instruction: How well do the following statements describe your personality?

I see myself as someone who ...	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
...is reserved	(1)	(2)	(3)	(4)	(5)
...is generally trusting	(1)	(2)	(3)	(4)	(5)
...tends to be lazy	(1)	(2)	(3)	(4)	(5)
...is relaxed, handles stress well	(1)	(2)	(3)	(4)	(5)
...has few artistic interests	(1)	(2)	(3)	(4)	(5)
...is outgoing, sociable	(1)	(2)	(3)	(4)	(5)
...tends to find fault with others	(1)	(2)	(3)	(4)	(5)
...does a thorough job	(1)	(2)	(3)	(4)	(5)
...gets nervous easily	(1)	(2)	(3)	(4)	(5)
...has an active imagination	(1)	(2)	(3)	(4)	(5)

Context (In)dependence This questionnaire is taken from Gollwitzer et al. (2006). The following is an English translation of the original questionnaire in German. Agreement to an item is measured on a 6 point Likert scale from "does not apply at all" to "fully applies".

Context dependence

1. My attitudes and opinions are often determined by the circumstances.
2. My behavior often depends on the people I am spending time with at that moment.
3. My decisions often depend on the temporary circumstances.
4. I behave very differently with different people.
5. My self-image depends overall on how other people perceive me.

Context independence

1. Once I have made a choice, I do not like to change it afterwards.
2. My self-image stays the same regardless of what others say about me.
3. I advocate for my own opinion regardless of the person with whom I am interacting.
4. I am the same person in different situations.
5. My attitudes and opinions hardly change, regardless of what happens in my life.

Moral disengagement This questionnaire is taken from Bandura et al. (1996). We excluded the following categories: euphemistic language, attribution of blame and dehumanization, as they did not apply to our experimental framework. The following is an English translation of the version by Rothmund (unpublished), who validated the questionnaire in German. Agreement to an item was measured on a 6-point Likert scale from "do not agree at all" to "fully agree".

1. It is alright to beat someone who badmouths your family.
2. Arriving late is better than not coming at all.
3. It does not make sense to avoid flying to go on vacation for the sake of the environment, since everybody else does it as well.
4. It is okay to tell small lies because they don't really do any harm.
5. It is alright to lie to keep your friends out of trouble.
6. Given the million-dollar frauds of some mangers, one cannot be blamed for scrounging some office supplies.
7. It is not so bad to cheat on taxes, since everybody does it anyway.
8. One cannot be blamed for an offence, if he or she has been put under pressure by his or her friends.
9. Teasing someone does not really hurt them.
10. It is less bad to steal from the rich than from the poor.
11. A single person cannot be blamed for misbehaving, if everyone else does the same.
12. Managers cannot be blamed for layoffs, that is simply how business life works.
13. It is alright to leave some trash in the cinema hall, since it will be cleaned after the screenplay anyway.
14. The reason why poor people do not have money is that they are too lazy to work.

Moral identity This questionnaire was originally developed by Aquino and Reed (2002). We use the German version validated by Rothmund and Gollwitzer (unpublished) and modified the list of attributes in the instructions. The following is an English translation of the material we used. Agreement to an item is measured on a 6-point Likert scale from "do not agree at all" to "fully agree".

Instructions: Below is a list of character attributes that might describe a person. The person with these attributes could be you, but also someone else.

Fair, generous, sympathetic, nice, and benign.

Imagine a person displaying exactly these character attributes. Imagine how this person would think, feel, and act. Once you have a precise image of this person, try to answer following questions.

1. It would make me feel good to be a person who has these characteristics.
2. Being someone who has these characteristics is an important part of who I am.
3. I would be ashamed to be a person who has these characteristics.
4. Having these characteristics is not really important to me.
5. I strongly desire to have these characteristics.
6. I often wear clothes that identify me as having these characteristics.
7. The types of things I do in my spare time (e.g., hobbies) clearly identify me as having these characteristics.
8. The kinds of books and magazines that I read identify me as having these characteristics.
9. The fact that I have these characteristics is conveyed to others by my membership in certain organizations.
10. I am actively involved in activities that convey to others that I have these characteristics.

Feelings Directly after the dictator game decision, subjects also went through a series of stages meant to investigate potential mechanisms driving our treatment effects. We describe the questions in the order in which they were presented to participants.

1. General happiness and contentment.
2. Feelings with regard to dictator game choice: happiness, guilt, content, amusement, shame, pride and excitement.[39]

D.1.3. Instructions

Welcome to the experiment. Thank you for your participation in this experiment. Please read the instructions carefully. For your participation today you will receive 5 €. During the experiment you will have the possibility to earn further money. Your additional payment will depend on your choices, the choices of other participants, as well as random events. Additionally, you will receive the earnings from the online part of the experiment at the end of today's experiment. After the experiment there will be a short questionnaire.

Please avoid any communication with your neighbors during the experiment. Switch off your mobile phone and remove everything you do not need for the experiment from the table. If you have any questions, please raise your hand and we will come to answer your questions at your seat.

In this experiment, a participant decides in the role of **Participant A** how to distribute 10 € between himself and another randomly determined **Participant B**.

First, all participants decide **in the role** of **Participant A**. This means that you will decide how to distribute **10 €** between yourself and **Participant B**. You can allocate any amount between 0 € and 10 € in discrete intervals to Participant B. Participant B will receive this amount and you will receive the remaining amount. Your decisions will be kept anonymous and you will not know, neither during nor after the experiment, with which participant you interacted.

You will learn which role you have been assigned to only at the end of the experiment and after you have taken your decision. Half of the participants will be assigned the role of Participant A, while the other half of the participants will be assigned that of Participant B. That is, there are two possibilities:

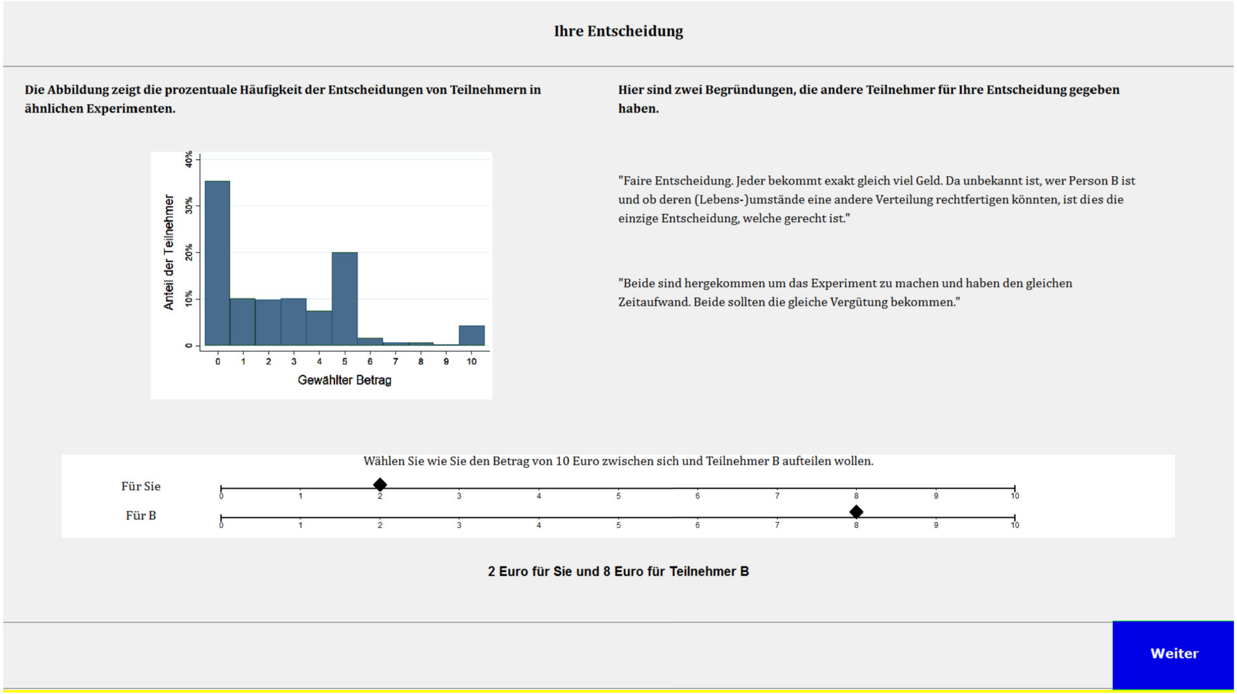
1. You are selected as Participant A. This means: Your decision will be implemented. You will be randomly assigned to someone in the role of Participant B. You will receive 10 €, minus the amount you have allocated to Participant B. Accordingly, Participant B will receive the amount you allocated him.
2. You are selected as Participant B. This means: Your decision will not be implemented. You will be randomly assigned to someone in the role of Participant A. You will receive an amount of money according to the decision of Participant A.

Since, at the time of making your decision, you do not know whether you will be selected as Participant A or Participant B, please take your decision carefully.

After the experiment, a short questionnaire will follow. Then, the experiment will be concluded. We kindly ask you to stay seated. We will call participants individually and pay them in private. Do you have further questions? Then, please raise your hand and we will come to answer your questions at your seat. Before the actual experiment starts, you will have to answer some questions of understanding.

[39] We also asked subjects to state their personal norm (Bašić and Verrina, 2021), i.e., how much they thought would be appropriate to give. However, since the measure was elicited after subjects had made their choice, we cannot exclude that it was used in a self-serving manner to further justify their choice. In fact, we find no variation between treatments and a high correlation with giving. For these reasons, we do not use this measure in our analysis.

D.1.4. Decision screen



Ihre Entscheidung

Die Abbildung zeigt die prozentuale Häufigkeit der Entscheidungen von Teilnehmern in ähnlichen Experimenten.

Hier sind zwei Begründungen, die andere Teilnehmer für Ihre Entscheidung gegeben haben.

"Faire Entscheidung. Jeder bekommt exakt gleich viel Geld. Da unbekannt ist, wer Person B ist und ob deren (Lebens-)umstände eine andere Verteilung rechtfertigen könnten, ist dies die einzige Entscheidung, welche gerecht ist."

"Beide sind hergekommen um das Experiment zu machen und haben den gleichen Zeitaufwand. Beide sollten die gleiche Vergütung bekommen."

Wählen Sie wie Sie den Betrag von 10 Euro zwischen sich und Teilnehmer B aufteilen wollen.

2 Euro für Sie und 8 Euro für Teilnehmer B

Fig. D.1. Dictator game decision screen. Note: The decision screen shows the empirical distribution of choices on the left. On the right side the two (positive or negative) narratives are listed. Below subjects take the dictator game decision.

D.1.5. Sessions (Table D.2)

Table D.2
Session overview - lab.

Session	Date (2018)	Treatment	Participants
1	May, 7	Baseline	22
2	May, 16	Baseline	24
3	May, 16	Baseline	28
4	May, 30	Positive	25
5	May, 30	Negative	22
6	May, 30	Positive	24
7	May, 30	Negative	26
8	June, 26	Positive	24
9	June, 26	Baseline	22
10	June, 26	Negative	25
11	June, 26	Negative	20
12	June, 26	Positive	18

D.2. Online experiment

As mentioned in Section 3.2, the online experiment can be considered as a rather demanding robustness check of the lab experiment for multiple reasons: different subject pool, less control and asynchronous decisions. One of our main concerns was to get high quality data. We took three main measures to achieve high data quality. First, our financial incentives are quite large by Prolific standards. They imply average hourly earnings of £19.61, compared to the recommended £7.50 ($9.60) and the minimum required payment of £5 ($6.50). Second, we screened out inattentive subjects through attention checks in both parts of the study. In total, 6% of subject were screened out in the attention check in the type elicitation experiment and 8% of in the dictator game experiment. Third, the experiment was kept simple and the questionnaire as short as possible in order to avoid confusion and frustration. Additionally, we sent subjects from the main experiment a message, once the payments to recipients were carried out.

In studies in which the experimenter provides information to subjects, demand effects may be a concern. In general, (De Quidt et al., 2018) find that these problems are less severe than commonly thought. In our specific case, it is unclear what direction the demand effect would have and how this could explain an effect of positive but not of negative narratives on giving. However, to address any remaining concern we ask subjects what they think the researchers expect to happen at the very end of the experiment. Only 13% of subjects mention the graph and only 13% of subjects in the treatment conditions mention the narratives as having anything to do with the research question. In addition, the subjects who correctly guess our hypothesis, i.e., they think that we expect more giving in the Positive and less in the Negative treatment, do not give significantly more or less compared to other subjects.

D.2.1. Narrative selection

Table D.3 shows positive and negative narratives (in their original version) along with their average convincingness rating. Numbers 1-2 were selected for the Positive condition and 3-4 for the Negative condition. 111 raters each rated 30 randomly selected narratives from a pool of 64 narratives which were taken from the first online Baseline session (all narratives that are linked to giving 0 or giving 5). A complete list of narratives is available from the authors upon request.

Table D.3
Narrative selection online.

Number	Positive Narratives	Convincingness
1	Even split is the fairest option. What if the person on the other end needs the dough as much as me? I can't deny them that.	6
2	I believe the other participant deserves as much as I do, regardless of one's capabilities or anything. I keep 5 points, and the other participant keeps same amount. It's fair.	6
	Negative Narratives	
3	I would normally share a reward equally for equal work, but since this other participant and I did not work on anything together, I don't feel that I owe any points to this random person.	4.7
4	I need the extra money, and because it is all anonymous I do not feel bad giving the other participant 0 points.	4.7

D.2.2. Additional psychological measures

We elicited a variation of the Comparison Orientation Measure by Gibbons and Buunk (1999) at the end of the type elicitation experiment. We selected the items which were relevant for our study and added four more (indicated in italics).

1. I always pay a lot of attention to how I do things compared with how others do things.
2. If I want to find out how well I have done something, I compare what I have done with how others have done.
3. I often compare how I am doing socially (e.g., social skills, popularity) with other people.
4. I am not the type of person who compares often with others (reversed).
5. I often compare myself with others with respect to what I have accomplished in life.
6. I often try to find out what others think who face similar problems as I face.
7. I always like to know what others in a similar situation would do.
8. I never consider my situation in life relative to that of other people.
9. *I do not want to look bad in comparison to others.*
10. *If someone has done something good, I try match his or her behavior.*
11. *I want to look better in comparison to others.*
12. *If someone has done something bad, I try distinguish myself from him or her behavior.*

D.2.3. Instructions

Welcome and thank you for taking part in this study. Please read the following instructions carefully.

This study consists of two parts. You have already completed the first part some days ago. This is the second part of the study. This part will take around 5 minutes of your time.

Today, on top of the standard payment for participation, you will earn a bonus of up to £2.50. The bonus payment depends solely on your decisions. You will learn more on this in the instructions that follow.

Note that you will be approved and receive the bonus and participation payments of this and the first part of the study only if you complete today's part!

Data & Consent

This study is entirely anonymous and you will not be asked to provide any personally identifying information. The responses you provide will only be used for research purposes and this data will be treated as highly confidential. No individual responses will be shared. By participating in this study, you consent to the data being used for this purpose. Your participation in this research is entirely voluntary. Please note that you have the right to withdraw consent or withdraw from the study at any point.

This study is conducted by researchers from the University of Mannheim (Germany) and the University of Lyon (France). If you have any question, please feel free to contact Adrian Hillenbrand at hillenbrand.requester@gmail.com.

(Attention check)

Instructions for the task:

Below are 3 statements that you can rate from ‘strongly disagree’ to ‘strongly agree’. Please consider your answers carefully. All this is not relevant for your task. The only purpose of this task is to exclude from the study all those participants who do not even read the instructions. This is necessary to ensure that only attentive respondents are considered in order to get interpretable answers.

Your task is to just pick the ‘strongly disagree’ answer for all 3 statements. This way we know you have read the instructions. Choosing any other option will lead to the direct exclusion from the study and any further payment.

Answer the questions according to the instructions above.

1. I am open to new experiences.
2. I make other people feel at ease.
3. I get stressed easily.

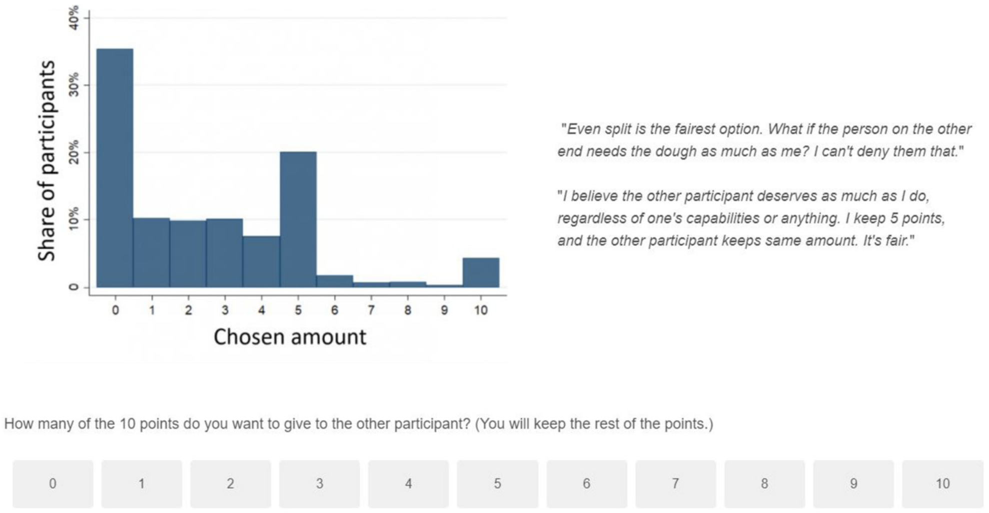
(Dictator Game) **Instructions for the task:**

In this task, you will decide how to split a given amount of money between yourself and another participant. You will be assigned 10 points, which are worth £2,50 (1 point = £0.25). You can decide how many points you want to give to the other participant, the rest you will keep for yourself.

Your choice will influence both your payment and that of another person. You will receive the amount of points converted in pounds that you decide to keep as a bonus payment at the end of the study.

The other participant will take part in an unrelated study of similar length to this one on Prolific. The participant will be provided with a description of this task but will not make a decision on his or her own. Your decision will be communicated to the other participant anonymously and it will influence the bonus payment of this participant. The other participant cannot influence this part of his or her payment and cannot influence your payment in any way. You will be informed via email once the payment to the other participant is performed.

D.2.4. Decision screen



"Even split is the fairest option. What if the person on the other end needs the dough as much as me? I can't deny them that."

"I believe the other participant deserves as much as I do, regardless of one's capabilities or anything. I keep 5 points, and the other participant keeps same amount. It's fair."

How many of the 10 points do you want to give to the other participant? (You will keep the rest of the points.)

Fig. D.2. Dictator game decision screen. Note: The decision screen shows the empirical distribution of choices on the left. On the right side the two (positive or negative) narratives are listed. Below subjects take the dictator game decision.

D.2.5. Sessions (Table D.4)

Table D.4
Session overview - online.

Session	Date (2021)	Treatment	Participants
1	June, 24	Baseline	90
2	July, 8	Positive, Negative	222
3	July, 14	Appropriateness	119
4	July, 16	Narratives categorization	99
5	September, 22	Baseline	47
6	September, 30	Positive, Negative	132
7	October, 6	Appropriateness	60
8	October, 14	Baseline	76
9	October, 28	Positive, Negative	123
10	November, 5	Narratives categorization	120
11	November, 5+11	Image concerns 1	101
12	December, 9	Image concerns 2	92

References

Akerlof, George A., Snower, Dennis J., 2016. Bread and bullets. J. Econ. Behav. Organ. 126, 58–71.

Akerlof, George A., Dickens, William T., 1982. The economic consequences of cognitive dissonance. Am. Econ. Rev. 72 (3), 307–319.

Andre, Peter, Haaland, Ingar, Roth, Christopher, Wohlfart, Johannes, 2021. Narratives about the Macroeconomy. CEBI Working Paper Series.

Andreoni, James, 1995. Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. Q. J. Econ. 110 (1), 1–21.

Andreoni, James, Rao, Justin M., 2011. The power of asking: how communication affects selfishness, empathy, and altruism. J. Public Econ. 95 (7–8), 513–520.

Aquino, Karl, Reed, I.I., 2002. The self-importance of moral identity. J. Pers. Soc. Psychol. 83 (6), 1423.

Balliet, Daniel, Parks, Craig, Joireman, Jeff, 2009. Social value orientation and cooperation in social dilemmas: a meta-analysis. Group Process. Intergroup Relat. 12 (4), 533–547.

Bandura, Albert, Barbaranelli, Claudio, Caprara, Gian Vittorio, Pastorelli, Concetta, 1996. Mechanisms of moral disengagement in the exercise of moral agency. J. Pers. Soc. Psychol. 71 (2), 364.

Bašić, Zvonimir, Verrina, Eugenio, 2021. Personal norms—and not only social norms—shape economic behavior. MPI Collective Goods Discussion Paper. 2020/25.

Bénabou, Roland, Tirole, Jean, 2006. Incentives and prosocial behavior. Am. Econ. Rev. 96 (5), 1652–1678.

Bénabou, Roland, Tirole, Jean, 2016. Mindful economics: the production, consumption, and value of beliefs. J. Econ. Perspect. 30 (3), 141–164.

Bénabou, Roland, Falk, Armin, Tirole, Jean, 2020. Narratives, Imperatives and Moral Persuasion.

Berg, Joyce, Dickhaut, John, McCabe, Kevin, 1995. Trust, reciprocity, and social history. Games Econ. Behav. 10 (1), 122–142.

Bicchieri, Cristina, Mercier, Hugo, 2013. Self-serving biases and public justifications in trust games. Synthese 190 (5), 909–922.

Böhm, Robert, Betsch, Cornelia, Korn, Lars, 2016. Selfish-rational non-vaccination: experimental evidence from an interactive vaccination game. J. Econ. Behav. Organ. 131, 183–195.

Bohnet, Iris, 1999. The sound of silence in prisoner's dilemma and dictator games. In: Economics as a Science of Human Behaviour. Springer, pp. 177–194.

Bott, Kristina M., Cappelen, Alexander W., Sørensen, Erik Ø, Tungodden, Bertil, 2019. You've got mail: a randomized field experiment on tax evasion. Manag. Sci.

Brañas-Garza, Pablo, 2007. Promoting helping behavior with framing in dictator games. J. Econ. Psychol. 28 (4), 477–486.

Bruner, Jerome, 1991. The narrative construction of reality. Crit. Inq. 18 (1), 1–21.

Bursztyn, Leonardo, Haaland, Ingar K., Rao, Aakaash, Roth, Christopher P., 2020. Disguising prejudice: Popular rationales as excuses for intolerant expression. Technical Report. National Bureau of Economic Research.

Cappelen, Alexander W., Cappelen, Cornelius, Tungodden, Bertil, 2018. Second-Best Fairness Under Limited Information: The Trade-Off Between False Positives and False Negatives. NHH Dept. of Economics Discussion Paper, vol. 18.

Cappelen, Alexander W., Moene, Karl O., Sørensen, Erik Ø, Tungodden, Bertil, 2013. Needs versus entitlements –- an international fairness experiment. J. Eur. Econ. Assoc. 11 (3), 574–598.

Cappelen, Alexander W., Halvorsen, Trond, Sørensen, Erik Ø, Tungodden, Bertil, 2017. Face-saving or fair-minded: what motivates moral behavior? J. Eur. Econ. Assoc. 15 (3), 540–557.

Carlson, Ryan W., André Maréchal, Michel, Oud, Bastiaan, Fehr, Ernst, Crockett, Molly J., 2020. Motivated misremembering of selfish decisions. Nat. Commun. 11 (1), 1–11.

Chance, Zoë, Norton, Michael I., Gino, Francesca, Ariely, Dan, 2011. Temporal view of the costs and benefits of self-deception. Proc. Natl. Acad. Sci. 108 (Supplement 3), 15655–15659.

Charness, Gary, Dufwenberg, Martin, 2006. Promises and partnership. Econometrica 74 (6), 1579–1601.

Collins, Sean M., Hamman, John R., Lightle, John P., 2018. Market interaction and pro-social behavior: an experimental study. South. Econ. J. 84 (3), 692–715.

Croson, Rachel, Marks, Melanie, 2001. The effect of recommended contributions in the voluntary provision of public goods. Econ. Inq. 39 (2), 238–249.

Dal Bó, Ernesto, Dal Bó, Pedro, 2014. "Do the right thing:" the effects of moral suasion on cooperation. J. Public Econ. 117, 28–38.

Dana, Jason, Weber, Roberto A., Xi Kuang, Jason, 2007. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Econ. Theory 33 (1), 67–80.

Danz, David, Vesterlund, Lise, Wilson, Alistair J., 2020. Belief elicitation: Limiting truth telling with information on incentives. Technical Report. National Bureau of Economic Research.

Ditto, Peter H., Pizarro, David A., Tannenbaum, David, 2009. Motivated moral reasoning. Psychol. Learn. Motiv. 50, 307–338.

Dreber, Anna, Ellingsen, Tore, Johannesson, Magnus, Rand, David G., 2013. Do people care about social context? Framing effects in dictator games. Exp. Econ. 16 (3), 349–371.

Engel, Christoph, 2011. Dictator games: a meta study. Exp. Econ. 14 (4), 583–610.

Epley, Nicholas, Gilovich, Thomas, 2016. The mechanics of motivated reasoning. J. Econ. Perspect. 30 (3), 133–140.

Exley, Christine L., 2015. Excusing selfishness in charitable giving: the role of risk. Rev. Econ. Stud. 83 (2), 587–628.

Falk, Armin, Becker, Anke, Dohmen, Thomas, Enke, Benjamin, Huffman, David, Sunde, Uwe, 2018. Global evidence on economic preferences. Q. J. Econ. 133 (4), 1645–1692.

Feiler, Lauren, 2014. Testing models of information avoidance with binary choice dictator games. J. Econ. Psychol. 45, 253–267.

Festinger, Leon, 1962. A Theory of Cognitive Dissonance, Vol. 2. Stanford University Press.

Fiedler, Susann, Glöckner, Andreas, Nicklisch, Andreas, Dickert, Stephan, 2013. Social value orientation and information search in social dilemmas: an eye-tracking analysis. Organ. Behav. Hum. Decis. Process. 120 (2), 272–284.

Fischbacher, Urs, 2007. z-tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10 (2), 171–178.

Foerster, Manuel, van der Weele, Joël J., 2021. Casting doubt: image concerns and the communication of social impact. Econ. J.

Galbiati, Roberto, Vertova, Pietro, 2008. Obligations and cooperative behaviour in public good games. Games Econ. Behav. 64 (1), 146–170.

Gibbons, Frederick X., Buunk, Bram P., 1999. Individual differences in social comparison: development of a scale of social comparison orientation. J. Pers. Soc. Psychol. 76 (1), 129.

Gino, Francesca, Norton, Michael I., Weber, Roberto A., 2016. Motivated Bayesians: feeling moral while acting egoistically. J. Econ. Perspect. 30 (3), 189–212.

Gino, Francesca, Ayal, Shahar, Ariely, Dan, 2009. Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. Psychol. Sci. 20 (3), 393–398.

Gino, Francesca, Ayal, Shahar, Ariely, Dan, 2013. Self-serving altruism? The lure of unethical actions that benefit others. J. Econ. Behav. Organ. 93, 285–292.

Goeree, Jacob K., Holt, Charles A., Laury, Susan K., 2002. Private costs and public benefits: unraveling the effects of altruism and noisy behavior. J. Public Econ. 83 (2), 255–276.

Gollwitzer, M., Schmidthals, K., Pöhlmann, C., 2006. Relationalitäts-Kontextabhängigkeits-Skala (RKS): Entwicklung und erste Ansätze zur Validierung. (Berichte aus der Arbeitsgruppe "Verantwortung, Gerechtigkeit, Moral" Nr. 161). Universität Trier, Trier.

Golman, Russell, Loewenstein, George, Moene, Karl Claire, Zarri, Luca, 2016. The preference for belief consonance. J. Econ. Perspect. 30 (3), 165–188.

Greiner, Ben, 2015. Subject pool recruitment procedures: organizing experiments with ORSEE. J. Econ. Sci. Assoc. 1 (1), 114–125.

Grossman, Zachary, Van Der Weele, Joël J., 2017. Self-image and willful ignorance in social decisions. J. Eur. Econ. Assoc. 15 (1), 173–217.

Haisley, Emily C., Weber, Roberto A., 2010. Self-serving interpretations of ambiguity in other-regarding behavior. Games Econ. Behav. 68 (2), 614–625.

Hamman, John R., Loewenstein, George, Weber, Roberto A., 2010. Self-interest through delegation: an additional rationale for the principal-agent relationship. Am. Econ. Rev. 100 (4), 1826–1846.

Hossain, Tanjim, Okui, Ryo, 2013. The binarized scoring rule. Rev. Econ. Stud. 80 (3), 984–1001.

Iriberri, Nagore, Rey-Biel, Pedro, 2011. The role of role uncertainty in modified dictator games. Exp. Econ. 14 (2), 160–180.

Kahneman, Daniel, Knetsch, Jack L., Thaler, Richard H., 1986. Fairness and the assumptions of economics. J. Bus., S285–S300.

Karlsson, Niklas, Loewenstein, George, McCafferty, Jane, et al., 2004. The economics of meaning. Nord. J. Polit. Econ. 30 (1), 61–75.

Konow, James, 2000. Fair shares: accountability and cognitive dissonance in allocation decisions. Am. Econ. Rev. 90 (4), 1072–1091.

Krämer, Florentin, Schmidt, Klaus M., Spann, Martin, Stich, Lucas, 2017. Delegating pricing power to customers: pay what you want or name your own price? J. Econ. Behav. Organ. 136, 125–140.

Krupka, Erin, Weber, Roberto A., 2009. The focusing and informational effects of norms on pro-social behavior. J. Econ. Psychol. 30 (3), 307–320.

Krupka, Erin L., Weber, Roberto A., 2013. Identifying social norms using coordination games: why does dictator game sharing vary? J. Eur. Econ. Assoc. 11 (3), 495–524.

Larson, Tara, Capra, C. Monica, 2009. Exploiting moral wiggle room: illusory preference for fairness? A comment. Judgm. Decis. Mak. 4 (6), 467.

Lazear, Edward P., Malmendier, Ulrike, Weber, Roberto A., 2012. Sorting in experiments with application to social preferences. Am. Econ. J. Appl. Econ. 4 (1), 136–163.

Matthey, Astrid, Regner, Tobias, 2011. Do I really want to know? A cognitive dissonance-based explanation of other-regarding behavior. Games 2 (1), 114–135.

Mazar, Nina, Amir, On, Ariely, Dan, 2008. The dishonesty of honest people: a theory of self-concept maintenance. J. Mark. Res. 45 (6), 633–644.

McAdams, Dan P., 1988. Power, Intimacy, and the Life Story: Personological Inquiries into Identity. Guilford Press.

Mohlin, Erik, Johannesson, Magnus, 2008. Communication: content or relationship? J. Econ. Behav. Organ. 65 (3–4), 409–419.

Murphy, Ryan, Ackermann, Kurt, Handgraaf, Michel, 2011. Measuring social value orientation. Judgm. Decis. Mak. 6 (8), 771–781.

Offerman, Theo, Sonnemans, Joep, Schram, Arthur, 1996. Value orientations, expectations and voluntary contributions in public goods. Econ. J., 817–845.

De Quidt, Jonathan, Haushofer, Johannes, Roth, Christopher, 2018. Measuring and bounding experimenter demand. Am. Econ. Rev. 108 (11), 3266–3302.

Rammstedt, Beatrice, John, Oliver P., 2007. Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. J. Res. Pers. 41 (1), 203–212.

Rodriguez-Lara, Ismael, Moreno-Garrido, Luis, 2012. Self-interest and fairness: self-serving choices of justice principles. Exp. Econ. 15 (1), 158–175.

Saucet, Charlotte, Villeval, Marie Claire, 2019. Motivated memory in dictator games. Games Econ. Behav. 117, 250–275.

Shalvi, Shaul, Gino, Francesca, Barkan, Rachel, Ayal, Shahar, 2015. Self-serving justifications: doing wrong and feeling moral. Curr. Dir. Psychol. Sci. 24 (2), 125–130.

Shalvi, Shaul, Dana, Jason, Handgraaf, Michel J.J., De Dreu, Carsten K.W., 2011. Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. Organ. Behav. Hum. Decis. Process. 115 (2), 181–190.

Shiller, Robert J., 2017. Narrative economics. Am. Econ. Rev. 107 (4), 967–1004.

van der Weele, Joël J., Kulisa, Julija, Kosfeld, Michael, Friebel, Guido, 2014. Resisting moral wiggle room: how robust is reciprocal behavior? Am. Econ. J. Microecon. 6 (3), 256–264.

Weisel, Ori, Zultan, Ro'i, 2016. Social motives in intergroup conflict: group identity and perceived target of threat. Eur. Econ. Rev. 90, 122–133.

Wiltermuth, Scott S., 2011. Cheating more when the spoils are split. Organ. Behav. Hum. Decis. Process. 115 (2), 157–168.

Xiao, Erte, 2017. Justification and conformity. J. Econ. Behav. Organ. 136, 15–28.