Discussion Papers of the Max Planck Institute for Research on Collective Goods Bonn 2018/16

The differential effect of narratives on prosocial behavior

Adrian Hillenbrand
Eugenio Verrina

MAX PLANCK SOCIETY

Discussion Papers of the
Max Planck Institute
for Research on Collective Goods

Bonn 2018/16

The differential effect of narratives on prosocial behavior

Adrian Hillenbrand / Eugenio Verrina

December 2018,

revised June 2020

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.coll.mpg.de

The differential effect of narratives on prosocial behavior

Adrian Hillenbrand and Eugenio Verrina*

June 8, 2020

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. Our main findings are that positive narratives increase giving of selfish types substantially, compared to a baseline with no narratives. Negative narratives, on the other hand, have a differential effect. Prosocial types decrease their giving, while selfish types give more than in the baseline. We argue and provide evidence in favor of the following interpretation of our results: narratives offer a benchmark for social comparison, on top of influencing perceptions of deservingness and appropriateness. Subjects are swayed by narratives and, at the same time, they compare themselves with the narrator.

Keywords: Prosocial behavior, narratives, social comparison, dictator game, SVO
JEL Classification: C91, D63, D64, D83, D91

*Hillenbrand: Max Planck Institute for Research on Collective Goods, 53113 Bonn, Germany (email: hillenbrand@coll.mpg.de). Verrina: Max Planck Institute for Research on Collective Goods, 53113 Bonn, Germany, and Cologne Graduate School in Management, Economics and Social Sciences, 50923 Cologne, Germany (email: verrina@coll.mpg.de). Financial support by the Max Planck Society is gratefully acknowledged. 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).

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 soaking up money from people who work hard. Besides, according to your colleague, the beggar will spend all the money you give him on alcohol and drugs; he deserves no consideration at all. Now imagine your colleague telling you instead that rising inequality is destroying our society and that the government does not do enough for people in need. He said we should all fight against the unfairness of this wicked capitalistic system. 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?

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).[1] In our opening illustration, the tension between a self-interested and a prosocial option 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 help explain fluctuations in markets (Shiller,

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

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[2] 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.[3] 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[4] behavior by leveraging the control of a laboratory experiment. In particular, we look at how positive and negative narratives that people use to justify their behavior influence the choice of others. Positive narratives, as defined by Bénabou et al. (2020), are arguments endorsing moral or prosocial behavior. Negative narratives, on the other hand, are arguments justifying immoral or selfish behavior. By controlling for the prosocial inclination of individuals, we analyze whether positive or negative narratives affect different types of individuals differently. 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 preferences greatly vary (Falk et al., 2018).

In our experiment, subjects play a dictator game where they decide how to share a given amount of money with another anonymous participant. In our two treatment conditions, they are shown either negative or positive narratives while making their choice. Narratives in the NEGATIVE condition

[2] 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.

[3] Foerster and van der Weele (2018a) work out a similar model where two agents with social image concerns can exchange signals about the social returns to an investment in a public good in a simultaneous 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. In a companion paper, Foerster and van der Weele (2018b) also test their model.

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

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.[5] We capitalize on arguments subjects use in previous experimental sessions for justifying their own choice to construct our treatments. 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). We thus look at how individuals who are more or less prosocial react to the narratives we present them. To that end, 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. Narratives, in our setting, are arguments targeting the perception of recipients' deservingness or of the appropriateness of giving. 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[6] types and should be stronger for prosocial types who receive a negative narrative and selfish types who receive a

[5] 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. 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 3.2), our hypotheses also hold in a social norms framework.

[6] 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.

positive narrative.

Our main results are that positive narratives increase giving, while negative narratives have a *differential effect* on different social types. In line with our predictions, types across the whole spectrum increase their giving in the POSITIVE condition, with selfish types displaying the largest effect. However, in the NEGATIVE condition, prosocial types decrease their giving, while selfish types increase their giving. This result is at odds with our hypotheses, according to which the same narrative cannot cause certain types to increase and other types to decrease giving.

We offer two potential explanations for this effect. According to the first, narratives - both positive and negative - enhance the salience of the moral decision, thus making it harder for subjects to behave selfishly. However, this explanation fails to account for part of our results, since it does not explain why negative narratives decrease the amount given by prosocial types. According to the second explanation, narratives provide a benchmark for social comparison. Subjects are, thus, induced to compare themselves with the narrator. Our social comparison explanation can account for the complete pattern of results including the differential effect: both positive and negative narratives increase giving of selfish types and negative narratives decrease giving of prosocial types. This explanation is also supported by additional results on the extensive and intensive margin of giving. Indeed, we find that positive narratives increase the probability of selfish types sharing the pie equally. On the other hand, negative narratives decrease the probability of selfish types giving nothing and do not increase it for prosocial types. Overall, our results can be explained by a desire to match the behavior of a prosocial narrator and to distinguish oneself from a selfish narrator by giving a bit more. This suggests that narratives may evoke a vivid comparison with the narrator beyond targeting perceptions of deservingness and appropriateness. We believe that capturing this motive can lead to important insights in prosocial behavior.

Our work sheds some first light on how narratives in the realm of prosocial and moral behavior work. From a practical standpoint, our results suggest that organizations and institutions can promote prosocial outcomes by confronting people with different narratives, positive or negative, depending on their predisposition. The evidence we present indicates that narratives

have the potential to increase prosocial behavior especially among those who would be less inclined to behave prosocially ex ante.

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, 2018a; Shiller, 2017; Akerlof and Snower, 2016) 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., 2011a; Wiltermuth, 2011; Rodriguez-Lara and Moreno-Garrido, 2012; Bicchieri and Mercier, 2013; Gino et al., 2013; Shalvi et al., 2015; Exley, 2015) 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 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. Dif-

ferently 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[7], 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.[8].

[7]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.

[8]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.

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. We find this to be indeed an important dimension to look at, since different types display not only quantitatively, but also qualitatively different reactions.

3 Experimental Design

3.1 Setup

Our experimental design consists of two main building blocks (see Figure 1), namely an online pre-study and a laboratory experiment. The laboratory experiment is subdivided in a modified dictator game and a questionnaire containing various ex-post measures. The online pre-study was conducted one week before the experiment.[9] The laboratory experiment was implemented in a between-subjects design with a BASELINE and two treatment conditions (POSITIVE and NEGATIVE), which varied only in the content of the narratives subjects saw. Below, we discuss the individual parts of the study in detail. Instructions for the laboratory experiment can be found in Appendix C.1.

[9]Subjects received the link to the pre-study one week before the experiment and had three days to complete it.

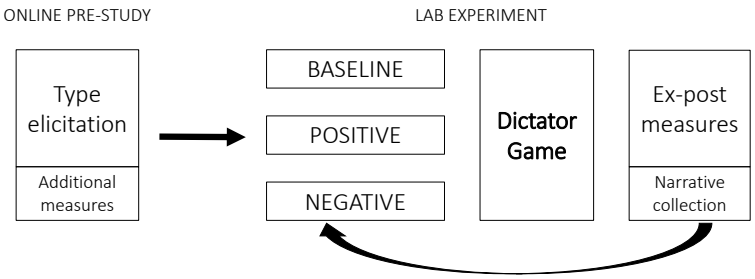

Figure 1: Experimental Design

Dictator game. The central part of our design is a simple dictator game (Kahneman et al., 1986). Dictators chose how to divide 10 € between themselves and an anonymous recipient (in intervals of 1 €). All subjects in the experiment decided under role uncertainty[10], 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.

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 (see Figure C1 in Appendix C.2). 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.[11] The figure displays the typical bimodal distribution with modes at 5 and

[10] 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.

[11] We used the following expression: "The figure shows the frequency of choices of participants in similar experiments in percentages."

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 their choices. These are our narratives (see Appendix C.3). In the POSITIVE condition, subjects saw two comments in support of the equal split (giving 5 €), while in the NEGATIVE condition they saw two comments justifying selfish behavior (giving 0 €). Subjects were (truthfully) told that these were explanations other participants had given for their choices in similar previous experiments.[12] 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 collection. After subjects had gone through all stages of the experiment, but before their final roles for the payment were revealed, they were given the opportunity, without any prior notice, to explain the reasoning behind their choice in the dictator game.[13] We used the explanations from the BASELINE condition to build the set of narratives subjects saw in the POSITIVE and NEGATIVE condition. Three independent raters, who were blind to the research question, evaluated the narratives along several dimensions. First, they were asked whether it was possible to understand what a subject had chosen in the dictator game from his or her comment and, if so, which was the most likely choice (0,1,2, etc.). Raters also evaluated how convincing they perceived the narrative to be (on a 7-point Likert

[12]We used the following expression: "Here are two explanations (*Begründungen*, in German), which other participants gave for their choice."

[13]The exact wording was the following. "You divided the money in the following way. You: €. Participant B: €. You can now explain ("*begründen*", in German) this decision for yourself." We asked subjects to stick to a maximum of two or three sentences and imposed a generous upper bound of 500 characters.

scale).[14]

We then selected the most convincing narratives in support of giving 0 € and in support of giving 5 € (using average ratings). We excluded narratives which were particularly long or repetitive. We selected four positive and four negative narratives. Each individual in the two treatment conditions saw two randomly selected narratives (at individual level). We take these steps, on the one hand, to prevent our results from depending on a single item and, on the other, to increase the probability of subjects indeed being treated by at least one narrative (see Appendix C.3 for the list of selected narratives).

Type elicitation. As mentioned above, the online pre-study was conducted one week prior to the laboratory experiment to avoid contamination across the two. The purpose of our online pre-study 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 below 22.45° are generally classified as individualists and those above as prosocials. Earnings in this task are determined by forming random pairs of subjects. One of the 6 choices is randomly selected and the choice of one of the two subjects in the pair is randomly implemented. 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

[14] Additionally, raters evaluated the narratives with regard to their creativity, profoundness, and honesty. We do not use these measures in this study.

capturing heterogeneity in reactions to our narrative manipulation.

We additionally elicit further psychological measures. 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) (for more details on these measures, see Appendix C.4). We use these measures (a) as controls in a robustness check in our regression analysis, and (b) to explore the role they play in explaining our treatment effect.

Ex-post measures. Directly after the dictator game decision, subjects 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.[15]

Procedures. The experiment was conducted at the DecisionLab of the Max Planck Institute for Research on Collective Goods in Bonn between May and June 2018.[16] The online experiment was conducted using Qualtrics, while the laboratory experiment was programmed in zTree (Fischbacher, 2007). Subjects were recruited via Orsee (Greiner, 2015). Before the start of the laboratory experiment subjects had to answer control questions to make sure they understood the experimental instructions correctly. 282 participants (64% female, average age 24.8 years)[17] took part in the experiment. For the analysis, we exclude 2 subjects who had not taken part in the online pre-study. Of the remaining 280 subjects, 96 subjects took part

[15]We also asked subjects to state their personal norm, 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.

[16]For an overview over all sessions, see Appendix C.5.

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

in the Baseline treatment, 91 in Positive, and 93 in Negative. All subjects received a show-up fee of 5 €, plus their earnings from the the online pre-study (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 €. The online pre-study lasted between 5 and 15 minutes, while the laboratory experiment took on average 40 minutes.

3.2 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.[18] 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 (e.g., Goeree et al., 2002, and see 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).[19] Similarly, decision makers care about the deservingness of the recipient(s)

[18] 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 and deservingness of the recipient of this action.

[19] 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.

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 lab or because roles were assigned by a random draw. 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 the perception, and hence the behavior, of different social types is influenced by negative and positive narratives. As mentioned above, the deservingness of a recipient and the appropriateness of giving are subject to uncertainty, and their perception can be influenced by narratives. This uncertainty leaves room for diverging perceptions.[20] In our setting, we call decision makers who perceive a recipient to be deserving or giving to be appropriate "prosocial" types, and the ones who believe the opposite "selfish" types.[21]

[20]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.

[21]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.

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, 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

Our dataset consists of 280 independent observations spread over three experimental conditions. In the first part of this section, we analyze the evidence regarding our main hypotheses. We then provide additional insights on the way our treatment conditions influence behavioral results by looking at whether subjects follow positive or negative narratives.

4.1 Main results

Subjects in the BASELINE condition give on average 2.76 €. According to Hypothesis 1, we should observe an increase in average giving in the POSITIVE condition and a decrease in the NEGATIVE condition. Figure 2 provides a visual representation of the aggregate results. In the POSITIVE condition, average giving increases to 3.23 €. This constitutes a 17% increase, in line with our first hypothesis. The difference, however, is only marginally significant (rank-sum test[22], $p = .093$). Average giving in the NEGATIVE condition (2.78 €) is virtually identical to average giving in the BASELINE condition (rank-sum test, $p = .908$).

However, the aggregate results on giving provide an incomplete picture of the data. As stated in Hypothesis 2, prosocial types should respond

[22] All tests are two-sided unless otherwise mentioned.

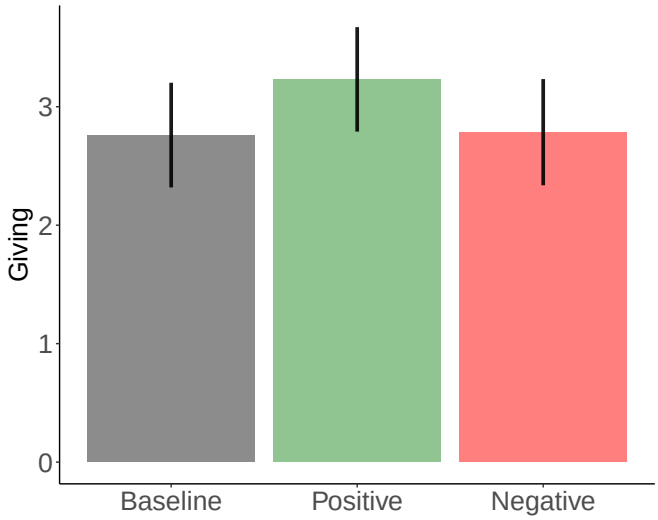

Figure 2: Average giving with 95%-confidence intervals.

more strongly to the NEGATIVE treatment condition and selfish types to the POSITIVE treatment condition. Although the effect should go in the same direction for all types.

Figure 3 displays the relationship between how much a subject gave in the dictator game and her social type. Giving is, as is typical in dictator games, bounded above at 5 € with only two subjects giving 6 € and many giving nothing at all. We use LOESS fitted lines to provide a better visualization of the data. The black solid line depicts the relationship between the social type and giving in BASELINE; the green dotted line represents our POSITIVE condition and the red dashed line our NEGATIVE condition. We observe the expected positive correlation between our social type measure and giving in the BASELINE condition. The steepness of the fitted line in the middle of the graph indicates that, in line with previous studies (see Engel, 2011), giving follows a bimodal distribution, with many subjects giving either half of their endowment or nothing at all.

To test how different types react to different narratives, we run a Tobit regression, 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 1). However, subjects' SVO-angles are not distributed uniformly (see Fig-

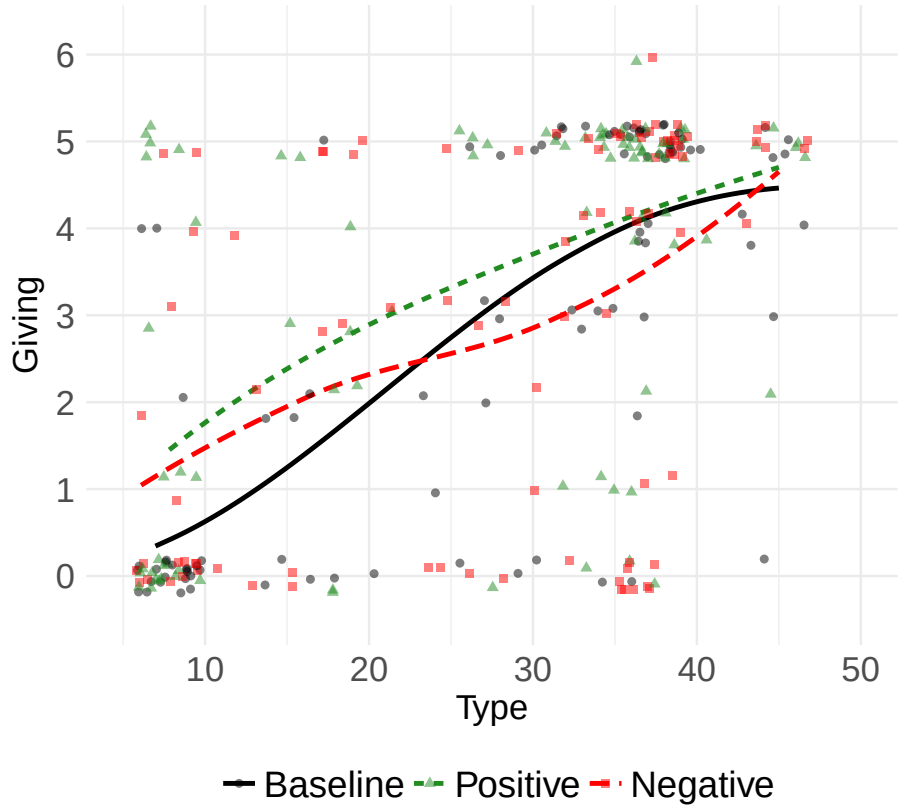

Figure 3: Giving on SVO. LOESS fitted lines.

Note: Data points are jittered. Black circles represent observations in the BASELINE, green triangles in the POSITIVE and red squares in the NEGATIVE treatment. For the ease of visualization, we removed social types below 5° and above 50°, which are rare (5 subjects) and not balanced across treatments.

ure 4).[23] The modal selfish type (60 subjects with an SVO angle of 7.82°) and the modal prosocial type (61 subjects with an SVO angle of 37.48°) make up 43% of all observations. Thus, we also look at them in isolation by reporting the estimated marginal effect and by performing separate tests to complement the regression analysis.[24] We discuss further robustness checks at the end of this section.

We first look at column (1) of Table 1, where we introduce our treat-

[23]Due to the construction of the measure specific SVO angles appear more frequently in the data (see Murphy et al., 2011).

[24]All our results go through when considering the definition of Murphy et al. (2011) for prosocials (SVO angle larger than 22.45°) and individualist (SVO angle lower than 22.45°).

dv: giving	(1)	(2)
POSITIVE	0.752**	2.852***
	(0.360)	(0.888)
NEGATIVE	0.125	2.698***
	(0.360)	(0.894)
Type	0.133***	0.189***
	(0.0116)	(0.0217)
POSITIVE × type		-0.0732**
		(0.0283)
NEGATIVE × type		-0.0900***
		(0.0285)
Constant	-1.382***	-3.015***
	(0.428)	(0.696)
Observations	280	280
Pseudo R^2	0.108	0.118

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table 1: Tobit regressions.

Note: Tobit regression with lower censoring at 0 (84 censored observations). The type measure corresponds to the SVO angle, POSITIVE and NEGATIVE conditions are included as dummies. We also include interaction terms between conditions and the SVO angle in column (2).

ment conditions as dummies and control for the social type of a subject. The POSITIVE condition has a strong positive and significant effect of on giving, confirming part of Hypothesis 1. The overall effect of the NEGATIVE condition is also positive, but small and not significant. Note that, as expected, the type measure is a clear predictor of giving: the higher the SVO angle of a subject, the more she gives.

In column (2) we add an interaction between subjects' social type and the treatment conditions. To interpret these results we plot the estimated marginal effects of our treatment conditions on giving compared to the BASELINE in Figure 4. This enables us to test Hypothesis 2.

We start with the POSITIVE condition (green dotted line), where we find a pattern in line with our hypothesis. We notice a strong positive effect for more selfish types, which fades out for more prosocial types. The estimated marginal effect for the modal selfish type corresponds to a

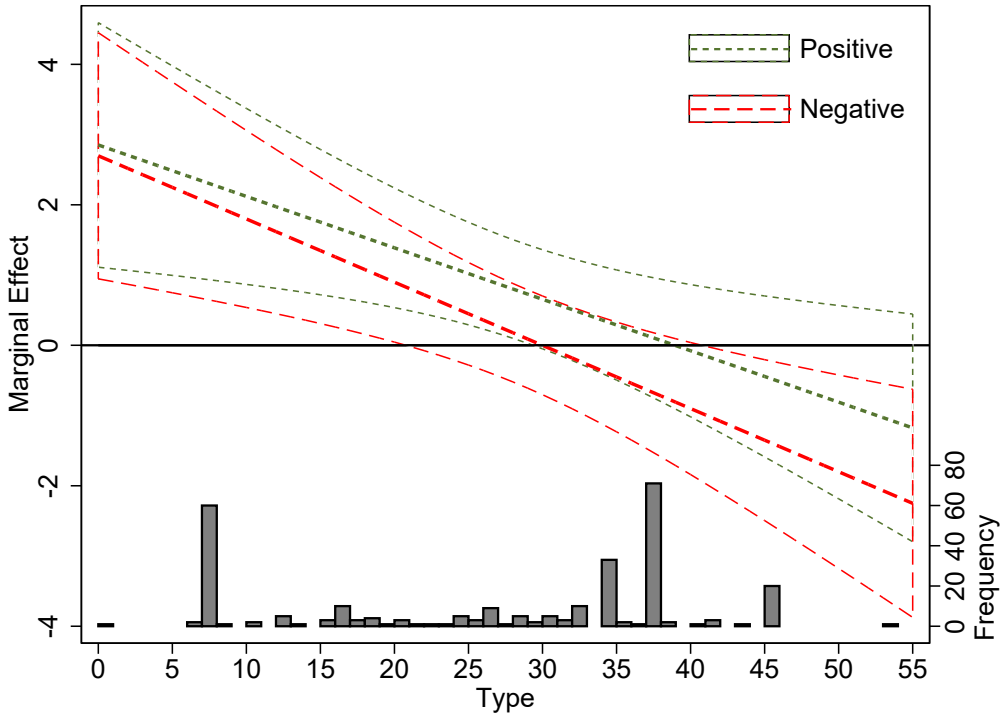

Figure 4: Marginal effects on types, Tobit.

Note: In the lower part of the graph, we plot the pooled distribution of types over all conditions. Numbers indicate the SVO-angle with higher angles indicating more prosociality. For the ease of visualization, types below 0° (3 subjects) are not displayed. Outer lines show 95% confidence intervals.

positive and significant difference of 2.28 € ($p = .001$) in giving, compared to the Baseline. Prosocial types, on the other hand, display no significant increase. This finding is corroborated by comparing giving in the Positive condition with the Baseline for the modal selfish (t-test, $N = 46$, $p = .028$) and prosocial types (t-test, $N = 39$, $p = .770$) in isolation.

Result 1 (Positive Narratives) *Positive narratives increase giving compared to the* Baseline *condition. This effect is driven by more selfish types.*

In the Negative condition (red dashed line), more selfish types increase their giving compared to the Baseline. The estimated marginal difference of 2 € ($p = .004$) for the modal selfish type is positive and significant. Note that this increase is indistinguishable from the one of the Positive condition. This is clearly not in line with our hypotheses. More

prosocial types, on the other hand, give less than in the BASELINE. The modal prosocial type decreases giving by an estimated marginal difference of 0.67 € ($p = .121$), which is not statistically significant. However, for more prosocial types (21 subjects with an SVO angle above 44°), the effect becomes negative and significant. These results are confirmed when restricting the analysis to only the modal selfish (t-test, $N = 37$, $p = .030$) and modal prosocial types (t-test, $N = 42$, $p = .016$), who increase and decrease giving, respectively.[25]

Result 2 (Negative Narratives) *Negative narratives have a differential effect: they decrease giving for more prosocial types and increase giving for selfish types compared to the* BASELINE.

We run further regressions to check the robustness of our results (see Appendix B). First, we include the additional psychological measures collected in the online pre-study and session dummies as controls in our Tobit model. We, then, check whether our results are robust to different specifications. We run a Tobit model with both lower and upper censoring. We also include a quadratic interaction term between our treatment conditions and the social type to capture potential nonlinearities. Finally, we compare our results with those of a standard OLS regression. Our results are robust to this additional analyses.[26]

4.2 Additional results: do people follow the narrative?

A natural question is whether narratives lead subjects to adhere to the behavioral prescription contained in them, i.e., either to share equally or keep everything for themselves. 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 two Probit regressions on the probability of giving either 5 or 0. The graphs in Figure 5 show the estimated

[25] Note that this is in line with the LOESS fit presented in Figure 3.

[26] We also perform our analysis using the Moral Identity Scale and the Moral Disengagement questionnaire as alternatives to the SVO angle in our main regression. Both have a strong and stable relationship with giving, but turn out to be irrelevant in explaining our treatment difference. Moreover, Context Dependence or Independence do not mediate our treatment effects. We discuss these results in Appendix B.1.

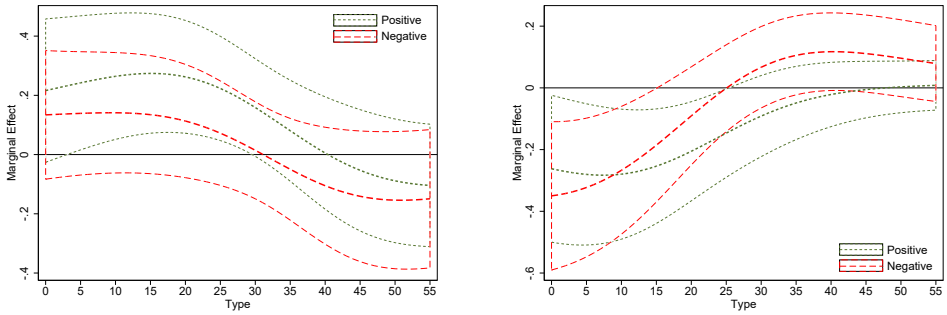

Figure 5: Marginal effects, Probit

Note: The dependent variable is a dummy for giving 5 € on the left and for giving 0 € on the right. Explanatory variables are: SVO angle, dummies for the POSITIVE and the NEGATIVE condition and interaction terms between treatment conditions and the SVO angle. Outer lines show 95 % confidence intervals. For the ease of visualization, subjects with an SVO angle below 0° (3 subjects) are not displayed.

marginal effects of the POSITIVE and NEGATIVE condition on different social types comapared to the BASELINE. We use the same specification as in our main regression in Table 1 column (2) (see Table B3 in Appendix B.2 for the full regression results).

There are three main observations to be made. First, the left graph in Figure 5 shows that the probability of giving an amount equal to 5 € in the POSITIVE condition increases for nearly all selfish types.[27] This translates into a 26% higher probability of giving 5 € for the modal selfish type in the POSITIVE condition (estimated marginal effect, $p = .022$). In the NEGATIVE condition, on the contrary, the increase in the probability of giving 5 € is smaller and statistically insignificant. The difference for the modal selfish type is 14% and not significant ($p = .178$). Second, the right graph in Figure 5 shows that both in the POSITIVE and the NEGATIVE condition the probability of selfish types giving 0 decreases substantially. This effect is observed across a wider range of types in the POSITIVE condition. The estimated marginal decrease in the probability of giving 0 € for the modal selfish type corresponds to 28% ($p = .012$) and 30% ($p = .007$) in the POSITIVE and NEGATIVE condition. Third, we find that, although

[27] The effect is particularly strong for the range of selfish types who are more frequent in our sample (those between an SVO angle of 5° and 25°).

more prosocial types give less in the NEGATIVE condition, this does not lead to a substantial increase in the probability of giving 0 €. The increase in probability for the modal prosocial type is moderate (11%) and only marginally significant ($p = .077$).

Result 3 *The* POSITIVE *condition increases the probability of giving* 5 € *for selfish types. Both treatment conditions decrease the probability of selfish types giving* 0 €.

We finally look at the effect of our treatment conditions on the ex-post measures of subjects' feelings (see Table B4 in Appendix B.3 for the regression analysis). We find no treatment effects on general happiness or contentment. Feelings of guilt and shame with regard to the choices made by subjects have, as one could expect, a strong and stable relation with the amount of giving: giving less increases these reported feelings. However, our treatment conditions do not increase or reduce guilt or shame about choices. 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.

Result 4 *Our treatment conditions do not directly change subjects' feelings towards their choice.*

5 Discussion and Conclusion

Our results provide insights into how narratives in favor of prosocial or selfish actions influence the behavior of different social types. Subjects in our experiment 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 work out two hypotheses from a theoretical framework on how narratives influence behavior via the perception of the appropriateness of an action or the deservingness of a recipient for different social types.

Subjects in the POSITIVE condition give more than subjects in the BASELINE condition. This increase is predominantly driven by selfish types (Result 1). On the other hand, narratives in the NEGATIVE condition have a differential effect (Result 2). Prosocial types in the NEGATIVE condition give less than in the BASELINE. This effect is reversed for selfish types, who give more in the NEGATIVE condition compared to the BASELINE, matching the giving of their peers in the POSITIVE condition. These results are only partly in line with the hypotheses derived from our theoretical framework. In particular, our hypotheses allow the effect of narratives to have different strength for different social types, but predict that all social types should move in the same direction. This suggests that narratives have an effect beyond that of arguing in favor or against the appropriateness of a certain action, as we describe below.

The differential effect of narratives resonates well with other research showing that different social types process information differently (Fiedler et al., 2013) and have a different representation of moral dilemmas (Van Lange et al., 1990; Liebrand et al., 1986). This suggests that our manipulation of positive and negative narratives could, indeed, affect prosocial and selfish types differently. We suggest two potential explanations for our results: one based on the argument that narratives enhance the moral saliency of the decision and another one based on a social comparison motive.

According to the first explanation, as pointed out above, the more selfish individuals might disregard the consequences of their actions and of the presence of a norm in their "ordinary" decision process. They could genuinely not know or ignore it. In both cases, the mere presence of a narrative, regardless of its content, could make the moral nature of the situation and, hence, the norm more salient, leading selfish individuals to give more. This conjecture is in line with a study 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. 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 account, however, does not explain why prosocial types decrease their giving when faced with a negative narrative, since the norm should be salient for them as well.

Our second explanation based on social comparison, instead, can account for the whole pattern of our results. If subjects care about how they fare in the comparison with others, the content of the narrative could serve as a social benchmark. In particular, narratives in the NEGATIVE condition would represent a very low reference point. Giving at least something after facing a negative narrative provides a low-cost opportunity for a selfish type to distinguish herself from the narrator. At the same time, prosocial types are led to give less by the negative narrative, but still care about faring well in the comparison with the narrator. In the POSITIVE condition, on the other hand, the social benchmark is set very high. For a subject not to look bad in this comparison she has to match the giving of the narrator. Taken together this would mean that subjects want to distinguish themselves from a selfish narrator and imitate a prosocial narrator. In Appendix A.1, we extend our model by including a social comparison component and provide a specification that can rationalize our results.

The additional results we obtain in Section 4.2 from our Probit regressions with either the equal split or giving nothing as dependent variables (Result 3) further support the social comparison explanation. The observed increase of equal splits in the POSITIVE condition suggests that at least some subjects wanted to avoid the negative comparison with the prosocial narrator and matched her giving. In the NEGATIVE condition, the probability of giving nothing decreases for selfish types and does not increase for prosocial types, implying that subjects were driven by a desire to differentiate themselves from the selfish narrator at least marginally. This behavior is in line with the phenomenon of partial lying (Fischbacher and Föllmi-Heusi, 2013) or ethical maneuvering (Mazar et al., 2008; Shalvi et al., 2011b), which is consistently found in the experimental literature on lying and cheating. Subjects often do not lie to the full extent, in order to avoid being unequivocally identified as liars or cheaters. This motivation is very similar to that of prosocial subjects in our experiment who decrease their giving, but do not go all the way to giving nothing at all.

Note that a social comparison explanation does not contradict the above point that narratives heighten the normative salience of the decision. Neither does it go against the evidence cited to support that explanation. Far from it, we in fact argue that narratives evoke a salient, vivid benchmark subjects compare themselves with. This account is supported by psychological theories which emphasize the importance of social comparison for people's self-evaluation (see the seminal work by Festinger, 1954; Suls and Wheeler, 2013, for and overview) and its crucial role for normative behavior (Cialdini et al., 1990, 2006). Social comparison has found fertile ground in economics as well and has sparked research in many different areas from energy and water conservation behavior (Allcott and Rogers, 2014; Ferraro and Price, 2013), to public good provision (Shang and Croson, 2009), charitable giving (Frey and Meier, 2004), all the way to retirement savings decisions (Beshears et al., 2015).

Importantly, our study was not designed to specifically test the social comparison mechanism. Psychological theories emphasize that for a comparison to be meaningful for an individual, she has to feel close to the person she is comparing herself with (Tesser, 1985). In other words, the comparison has to be self-relevant. In our experiment, we did not manipulate the relevance of the comparison with the narrator. This could be done, e.g., by choosing narrators that either belong to the same social category of the subject or to a different one. Since we use a student sample and subjects in our sample are used to face other students in these experiment, there are good reasons to believe that the comparison was relevant for them.

Our work advances the understanding of the determinants of prosocial and moral behavior by providing insights into how narratives - which permeate people's life - work. Our findings suggest that narratives sway subjects while, at the same time, serving as a benchmark for social comparison. Arguments in favor of selfish or prosocial behavior seem to evoke a concrete normative dilemma in subjects' mind. To be or not to be like the narrator? How will I fare compared to her? Subjects react to this vivid image by adhering to the narrative of a prosocial narrator and wanting to distinguish themselves from a selfish narrator. Certainly, more research is needed to understand how exactly this process works.

Our results also have relevant implications for institutions and organi-

zations who can use narratives to promote prosocial behavior, especially amongst the people who would be less inclined to act so ex ante. This can be achieved by confronting people with different narratives, positive or negative, depending on their predisposition. In the setting we study, sharing the money equally represents a clear norm of behavior. 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.

References

Akerlof, George A and Dennis J Snower, "Bread and bullets," *Journal of Economic Behavior & Organization*, 2016, *126*, 58–71.

_ **and William T Dickens**, "The economic consequences of cognitive dissonance," *The American Economic Review*, 1982, *72* (3), 307–319.

Allcott, Hunt and Todd Rogers, "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation," *American Economic Review*, 2014, *104* (10), 3003–37.

Andreoni, James, "Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments," *The Quarterly Journal of Economics*, 1995, *110* (1), 1–21.

_ **and Justin M Rao**, "The power of asking: How communication affects selfishness, empathy, and altruism," *Journal of Public Economics*, 2011, *95* (7-8), 513–520.

Aquino, Karl and I.I. Reed, "The self-importance of moral identity.," *Journal of Personality and Social Psychology*, 2002, *83* (6), 1423.

Balliet, Daniel, Craig Parks, and Jeff Joireman, "Social value orientation and cooperation in social dilemmas: A meta-analysis," *Group Processes & Intergroup Relations*, 2009, *12* (4), 533–547.

Bandura, Albert, Claudio Barbaranelli, Gian Vittorio Caprara, and Concetta Pastorelli, "Mechanisms of moral disengagement in the exercise of moral agency.," *Journal of Personality and Social Psychology*, 1996, *71* (2), 364.

Bénabou, Roland and Jean Tirole, "Incentives and prosocial behavior," *American Economic Review*, 2006, *96* (5), 1652–1678.

_ **and** _ , "Mindful economics: The production, consumption, and value of beliefs," *Journal of Economic Perspectives*, 2016, *30* (3), 141–64.

_ **, Armin Falk, and Jean Tirole**, "Narratives, Imperatives and Moral Persuasion," 2020.

Beshears, John, James J Choi, David Laibson, Brigitte C Madrian, and Katherine L Milkman, "The effect of providing peer information on retirement savings decisions," *The Journal of finance*, 2015, *70* (3), 1161–1201.

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

Bó, Ernesto Dal and Pedro Dal Bó, ""Do the right thing:" The effects of moral suasion on cooperation," *Journal of Public Economics*, 2014, *117*, 28–38.

Böhm, Robert, Cornelia Betsch, and Lars Korn, "Selfish-rational non-vaccination: experimental evidence from an interactive vaccination game," *Journal of Economic Behavior & Organization*, 2016, *131*, 183–195.

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

Bott, Kristina M, Alexander W Cappelen, Erik Ø Sørensen, and Bertil Tungodden, "You've got mail: A randomized field experiment on tax evasion," *Management Science*, 2019.

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

Bruner, Jerome, "The narrative construction of reality," *Critical Inquiry*, 1991, *18* (1), 1–21.

Cappelen, Alexander W, Cornelius Cappelen, and Bertil Tungodden, "Second-best fairness under Limited information: The trade-off between false positives and false negatives," *NHH Dept. of Economics Discussion Paper*, 2018, (18).

_ , Karl O Moene, Erik Ø Sørensen, and Bertil Tungodden, "Needs versus entitlements—an international fairness experiment," *Journal of the European Economic Association*, 2013, *11* (3), 574–598.

_ , **Trond Halvorsen, Erik Ø Sørensen, and Bertil Tungodden**, "Face-saving or fair-minded: What motivates moral behavior?," *Journal of the European Economic Association*, 2017, *15* (3), 540–557.

Carlson, Ryan W, Michel André Maréchal, Bastiaan Oud, Ernst Fehr, and Molly J Crockett, "Motivated misremembering of selfish decisions," *Nature Communications*, 2020, *11* (1), 1–11.

Chance, Zoë, Michael I Norton, Francesca Gino, and Dan Ariely, "Temporal view of the costs and benefits of self-deception," *Proceedings of the National Academy of Sciences*, 2011, *108* (Supplement 3), 15655–15659.

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

Cialdini, Robert B, Linda J Demaine, Brad J Sagarin, Daniel W Barrett, Kelton Rhoads, and Patricia L Winter, "Managing social norms for persuasive impact," *Social influence*, 2006, *1* (1), 3–15.

_ , **Raymond R Reno, and Carl A Kallgren**, "A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places.," *Journal of personality and social psychology*, 1990, *58* (6), 1015.

Croson, Rachel and Melanie Marks, "The effect of recommended contributions in the voluntary provision of public goods," *Economic Inquiry*, 2001, *39* (2), 238–249.

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

Ditto, Peter H, David A Pizarro, and David Tannenbaum, "Motivated moral reasoning," *Psychology of Learning and Motivation*, 2009, *50*, 307–338.

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

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

Epley, Nicholas and Thomas Gilovich, "The mechanics of motivated reasoning," *Journal of Economic Perspectives*, 2016, *30* (3), 133–40.

Exley, Christine L, "Excusing selfishness in charitable giving: The role of risk," *The Review of Economic Studies*, 2015, *83* (2), 587–628.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global evidence on economic preferences," *The Quarterly Journal of Economics*, 2018, *133* (4), 1645–1692.

Feiler, Lauren, "Testing models of information avoidance with binary choice dictator games," *Journal of Economic Psychology*, 2014, *45*, 253–267.

Ferraro, Paul J and Michael K Price, "Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment," *Review of Economics and Statistics*, 2013, *95* (1), 64–73.

Festinger, Leon, "A theory of social comparison processes," *Human Relations*, 1954, *7* (2), 117–140.

_ , *A theory of cognitive dissonance*, Vol. 2, Stanford university press, 1962.

Fiedler, Susann, Andreas Glöckner, Andreas Nicklisch, and Stephan Dickert, "Social value orientation and information search in social dilemmas: An eye-tracking analysis," *Organizational Behavior and Human Decision Processes*, 2013, *120* (2), 272–284.

Fischbacher, Urs, "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 2007, *10* (2), 171–178.

_ **and Franziska Föllmi-Heusi**, "Lies in disguise—an experimental study on cheating," *Journal of the European Economic Association*, 2013, *11* (3), 525–547.

Foerster, Manuel and Joel J van der Weele, "Denial and Alarmism in Collective Action Problems," 2018.

_ **and** _ , "Persuasion, justification and the communication of social impact," 2018.

Frey, Bruno S and Stephan Meier, "Social comparisons and pro-social behavior: Testing" conditional cooperation" in a field experiment," *American Economic Review*, 2004, *94* (5), 1717–1722.

Galbiati, Roberto and Pietro Vertova, "Obligations and cooperative behaviour in public good games," *Games and Economic Behavior*, 2008, *64* (1), 146–170.

Gino, Francesca, Michael I Norton, and Roberto A Weber, "Motivated Bayesians: Feeling moral while acting egoistically," *Journal of Economic Perspectives*, 2016, *30* (3), 189–212.

_ **, Shahar Ayal, and Dan Ariely**, "Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel," *Psychological Science*, 2009, *20* (3), 393–398.

_ **,** _ **, and** _ , "Self-serving altruism? The lure of unethical actions that benefit others," *Journal of Economic Behavior & Organization*, 2013, *93*, 285–292.

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

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

Golman, Russell, George Loewenstein, Karl Ove Moene, and Luca Zarri, "The preference for belief consonance," *Journal of Economic Perspectives*, 2016, *30* (3), 165–88.

Greiner, Ben, "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 2015, *1* (1), 114–125.

Grossman, Zachary and Joel J Van Der Weele, "Self-image and willful ignorance in social decisions," *Journal of the European Economic Association*, 2017, *15* (1), 173–217.

Haisley, Emily C and Roberto A Weber, "Self-serving interpretations of ambiguity in other-regarding behavior," *Games and Economic Behavior*, 2010, *68* (2), 614–625.

Hamman, John R, George Loewenstein, and Roberto A Weber, "Self-interest through delegation: An additional rationale for the principal-agent relationship," *American Economic Review*, 2010, *100* (4), 1826–1846.

Iriberri, Nagore and Pedro Rey-Biel, "The role of role uncertainty in modified dictator games," *Experimental Economics*, 2011, *14* (2), 160–180.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler, "Fairness and the assumptions of economics," *Journal of Business*, 1986, pp. S285–S300.

Karlsson, Niklas, George Loewenstein, Jane McCafferty et al., "The economics of meaning," *Nordic Journal of Political Economy*, 2004, *30* (1), 61–75.

Konow, James, "Fair shares: Accountability and cognitive dissonance in allocation decisions," *American Economic Review*, 2000, *90* (4), 1072–1091.

Krämer, Florentin, Klaus M Schmidt, Martin Spann, and Lucas Stich, "Delegating pricing power to customers: Pay what you want or name your own price?," *Journal of Economic Behavior & Organization*, 2017, *136*, 125–140.

Krupka, Erin and Roberto A Weber, "The focusing and informational effects of norms on pro-social behavior," *Journal of Economic Psychology*, 2009, *30* (3), 307–320.

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

Lange, Paul AM Van, Wim BG Liebrand, and D Michael Kuhlman, "Causal attribution of choice behavior in three N-person prisoner's dilemmas," *Journal of Experimental Social Psychology*, 1990, *26* (1), 34–48.

Larson, Tara and C Monica Capra, "Exploiting moral wiggle room: Illusory preference for fairness? A comment," *Judgment and Decision Making*, 2009, *4* (6), 467.

Lazear, Edward P, Ulrike Malmendier, and Roberto A Weber, "Sorting in experiments with application to social preferences," *American Economic Journal: Applied Economics*, 2012, *4* (1), 136–63.

Liebrand, Wim BG, Ronald WTL Jansen, Victor M Rijken, and Cor JM Suhre, "Might over morality: Social values and the perception of other players in experimental games," *Journal of Experimental Social Psychology*, 1986, *22* (3), 203–215.

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

Mazar, Nina, On Amir, and Dan Ariely, "The dishonesty of honest people: A theory of self-concept maintenance," *Journal of Marketing Research*, 2008, *45* (6), 633–644.

McAdams, Dan P, *Power, intimacy, and the life story: Personological inquiries into identity*, Guilford Press, 1988.

Mohlin, Erik and Magnus Johannesson, "Communication: Content or relationship?," *Journal of Economic Behavior & Organization*, 2008, *65* (3-4), 409–419.

Murphy, Ryan, Kurt Ackermann, and Michel Handgraaf, "Measuring social value orientation," *Judgment and Decision Making*, 2011, *6* (8), 771–781.

Offerman, Theo, Joep Sonnemans, and Arthur Schram, "Value orientations, expectations and voluntary contributions in public goods," *The Economic Journal*, 1996, pp. 817–845.

Rammstedt, Beatrice and Oliver P John, "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German," *Journal of Research in Personality*, 2007, *41* (1), 203–212.

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

Saucet, Charlotte and Marie Claire Villeval, "Motivated memory in dictator games," *Games and Economic Behavior*, 2019, *117*, 250–275.

Shalvi, Shaul, Francesca Gino, Rachel Barkan, and Shahar Ayal, "Self-serving justifications: Doing wrong and feeling moral," *Current Directions in Psychological Science*, 2015, *24* (2), 125–130.

_ , Jason Dana, Michel JJ Handgraaf, and Carsten KW De Dreu, "Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior," *Organizational Behavior and Human Decision Processes*, 2011, *115* (2), 181–190.

_ , Michel JJ Handgraaf, and Carsten KW De Dreu, "Ethical manoeuvring: Why people avoid both major and minor lies," *British Journal of Management*, 2011, *22*, S16–S27.

Shang, Jen and Rachel Croson, "A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods," *The economic journal*, 2009, *119* (540), 1422–1439.

Shiller, Robert J, "Narrative economics," *American Economic Review*, 2017, *107* (4), 967–1004.

Suls, Jerry and Ladd Wheeler, *Handbook of social comparison: Theory and research*, Springer Science & Business Media, 2013.

Tesser, Abraham, "Toward a self-evaluation maintenance model of social behavior.," 1985.

van der Weele, Joël J, Julija Kulisa, Michael Kosfeld, and Guido Friebel, "Resisting moral wiggle room: how robust is reciprocal behavior?," *American Economic Journal: Microeconomics*, 2014, *6* (3), 256–64.

Weisel, Ori and Roi Zultan, "Social motives in intergroup conflict: Group identity and perceived target of threat," *European Economic Review*, 2016, *90*, 122–133.

Wiltermuth, Scott S, "Cheating more when the spoils are split," *Organizational Behavior and Human Decision Processes*, 2011, *115* (2), 157–168.

Xiao, Erte, "Justification and conformity," *Journal of Economic Behavior & Organization*, 2017, *136*, 15–28.

A Appendix: theoretical framework

This section complements the "Behavioral Predictions" in the main text (Section 3.2) 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. Finally, we provide an extension of the model with an additional component that captures our social comparison explanation.

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.[28] 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 concavity 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 3.2). 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, by

[28] 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.

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.[29]

As an example, let us assume a signal structure as in Figure A1. If there is no externality $E = 0$, with probability $1 \geq c > \frac{1}{2}$ the correct signal, i.e. the negative narrative, is sent, and with $1 - c$ the signal is wrong, i.e. the narrative is positive. The situation is reversed with a high externality ($E = 1$).

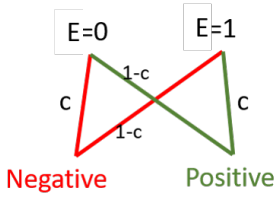

Figure A1: Examplary signal structure

The posterior given a positive or negative signal is calculated as follows (with e being the prior probability of $E = 1$). Figure A2 provides a graphical representation.

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

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

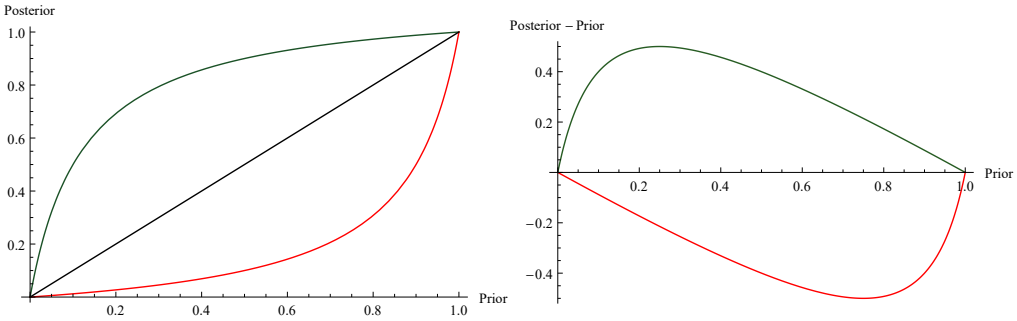

Figure A2: Posterior for given signal

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 $c = 0.9$. The black line on the left is the 45-degree line representing the case with no signal or no updating.

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.[30] 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

[30] 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.

way, i.e., choose to give the same amount, if they held the same belief. Modelling 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 (Figure A2).

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.*

A.1 Extension: Social Comparison

We provide an extension of our model and analyse the optimal giving behavior for a specification which captures our main results as well as our additional ones. Note that the goal is not to offer a general solution to the analytical problem here, but rather to show that the addition of a social comparison component can explain our findings.

The main idea is that narratives, on top of acting as a signal, provide a benchmark for social comparison. We introduce a social comparison component to the utility function which captures this intuition. DMs gain from giving more than the narrator but this gain is decreasing for larger amounts, i.e., gains are concave in the positive difference between giving and the amount advocated for by the narrator. Conversely, giving a little less than the narrator leads to a large loss which marginally decreases for lower giving, i.e., it is convex in the difference of giving and the narrator's giving. The following specification reflects this and Figure A3 shows a potential social comparison function:

$$S(g,n) = \begin{cases} \mu(g-n)^{\alpha} & \text{if } g \geq n \\ -\mu(n-g)^{\alpha} & \text{if } g < n, \end{cases} \tag{2}$$

with $\alpha < 1$, where n determines the narrator's amount of giving and μ the weighting of the social comparison.

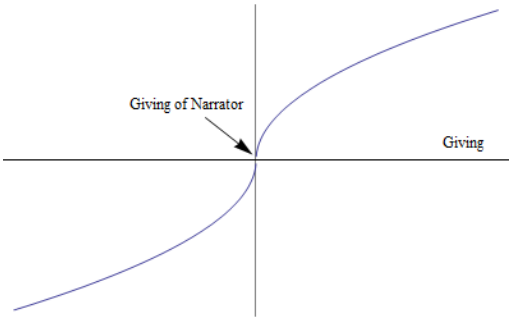

Figure A3: Social comparison function

Note: An example for a social comparison function with $\alpha = 0.5$.

We define the utility function of DMs as

$$U(g, e) = v(g, e) + S(g, n) - c(g), \tag{3}$$

where $S(g, n)$ describes a social comparison function as above. Importantly, the social comparison part is evoked by a narrative and disappears if there is no narrative (as in our BASELINE treatment).

The general solution of the above problem is not straightforward and might depend on the specifications of the value and social comparison functions. For tractability, we use a specification with a linear increasing giving function. This will lead to a step-function of giving for the POSITIVE as well as for the BASELINE treatment. We take the following specification:

$$U(g, e) = \begin{cases} 2eg + \mu(g - n)^{\frac{1}{2}} - g & \text{if } g \geq n \\ 2eg - \mu(n - g)^{\frac{1}{2}} - g & \text{if } g < n \end{cases} \tag{4}$$

with $g \in [0, 1]$ and e being the belief about the presence of the externality which can be influenced by a narrative as above.

In the example, we define the amount given by the narrator of a positive narrative as $n = 1$ (this reflects the natural, fair upper bound of giving 5 € in our experiment), and that of a narrator of a negative narrative as $n = 0$.

The resulting optimal giving functions in the three treatments are displayed in Figure A4 and formally presented below (e_{pos} and e_{neg} reflect the

posterior beliefs about the presence of an externality in the positive and negative narrative treatment, respectively).

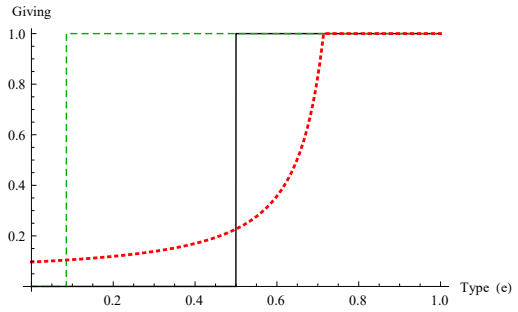

Figure A4: Predicted giving behavior

Note: The figure shows the predicted giving functions for the above specification in the BASELINE (black), NEGATIVE (red, dotted), and POSITIVE condition (green, dashed). Example parameters $\mu = 0.39$, $c = 0.83$.

$$g^*(e,\mu)_{Baseline} = \begin{cases} 0 & \text{if } e < \frac{1}{2} \\ 1 & \text{if } e \geq \frac{1}{2} \end{cases} \tag{5}$$

$$g^*(e,\mu)_{Positive} = \begin{cases} 0 & \text{if } e_{pos}(c,e) < \frac{1-\mu}{2} \\ 1 & \text{if } e_{pos}(c,e) \geq \frac{1-\mu}{2} \end{cases} \tag{6}$$

$$g^*(e,\mu)_{Negative} = \min((\frac{\mu}{1-2e_{neg}})^2, 1) \tag{7}$$

The resulting predicted behavior according to the model shares key characteristics with our experimental results (see Figure 3 in Section 4). First, giving is higher in POSITIVE compared to BASELINE. Second, there is a differential effect for NEGATIVE narratives with prosocial types decreasing their giving and selfish types increasing their giving. Importantly, this example also captures the increase in equal splits, which is the action the narrator advocates for in POSITIVE, and the decrease of subjects giving nothing in NEGATIVE.

B Appendix: Robustness checks

In Table B1 we conduct multiple robustness checks. In the first column we control for the additional psychological measures.[31] In column 2, we impose both lower and upper censoring. For interpretability of the interactions, we plot marginal effects as in the main text (see Figure B1). Column 3 introduces a quadratic term for types and interactions with the treatment conditions (see Figure B2 for the marginal effects). We normalize our type measure for this specification (in the graph, we show the most frequent non-normalized types as references). The pattern described in Section 4 remains substantively the same for all these alternative specifications. In column 4, we run a standard OLS regression. Also in this case, results are comparable to those of our main regressions.

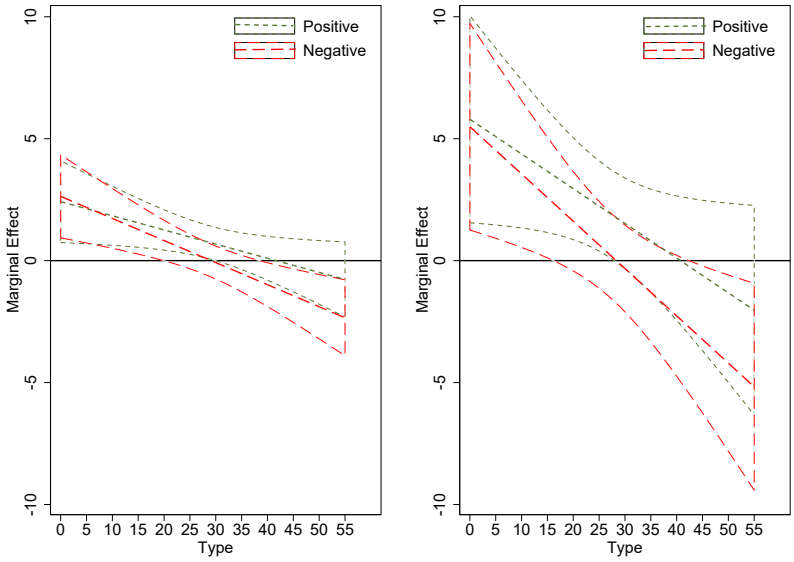

Figure B1: Marginal effects, Tobit.

Note: Tobit with lower censoring at 0 and controls on the left, Tobit with upper and lower censoring (5 and 0) on the right. Outer lines show 95 %-confidence intervals.

[31] 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.

dv: giving	Tobit controls	Tobit sessions	Tobit, upper and lower censoring	Tobit quadratic	OLS
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

* $p < .10$, ** $p < .05$, *** $p < .01$

Table B1: Robustness checks

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 conditions and types. Controls include Context Dependence, Context Independence, Moral Identity Scale, Moral Disengagement, and the 11-item, Big-5 questionnaire. Session includes session dummies. In the Tobit 84 observations are censored at giving 0 and 120 observations censored at giving of 5.

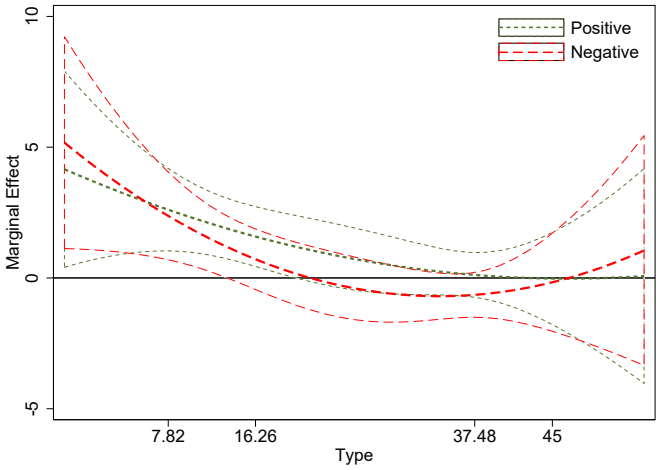

Figure B2: Marginal effects, Tobit

Note: Tobit with quadratic interaction term. Outer lines show 95 % confidence intervals

B.1 Analysis of additional psychological measures

In Table B2, we run the same analysis as in Section 4 using the additional psychological measures collected in the online pre-study. 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.

dv: giving	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

* $p < .10$, ** $p < .05$, *** $p < .01$

Table B2: Tobit regression, alternative measures

Note: Tobit regression 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 conditions and the type measure.

B.2 Probit regressions

	give 5	give 0
POSITIVE	1.559**	-0.975**
	(0.653)	(0.471)
NEGATIVE	1.020	-1.230***
	(0.694)	(0.457)
Type	0.0820***	-0.0825***
	(0.0167)	(0.0131)
POSITIVE x type	-0.0386*	0.0204
	(0.0198)	(0.0186)
NEGATIVE x type	-0.0328	0.0491***
	(0.0209)	(0.0166)
Constant	-2.705***	1.617***
	(0.568)	(0.352)
Observations	280	280
Pseudo R^2	0.213	0.275

Standard errors in parentheses

* $p < .10$, ** $p < .05$, *** $p < .01$

Table B3: Probit regressions, giving 5 and giving 0

Note: Probit regression. The dependent variable is giving 5 in the first column and 0 in the second column, The type measure corresponds to the SVO angle, POSITIVE and NEGATIVE conditions are included as dummies. We also include interaction terms between conditions and types.

B.3 Feelings

In Table B4, 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.

	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
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table B4: OLS regressions, feelings

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 conditions and types.

C Appendix: additional materials

This Appendix contains additional materials used for the experiment.

C.1 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.

Instructions

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.

C.2 Decision Screen

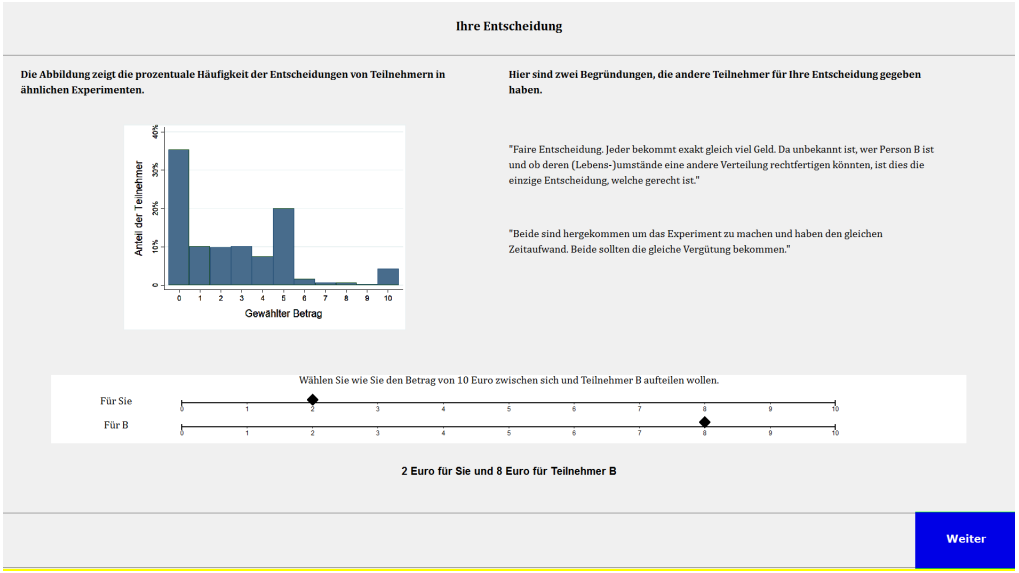

Figure C1: Dictator game decision screen

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.

C.3 Narrative Selection

The following table 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. Narratives were selected from all narratives of the first 3 sessions of the BASELINE condition, since the 4th session was run later to balance the number of participants in all conditions. More detailed information as well as the complete list of comments is available from the authors upon request.

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

C.4 Additional psychological measures

C.4.1 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)

C.4.2 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.

C.4.3 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.

C.4.4 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.

C.5 Sessions

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

Table C1: Session overview