DISCUSSION PAPER SERIES

Social Proximity and the Erosion of Norm Compliance

Cristina Bicchieri
Eugen Dimant
Simon Gächter
Daniele Nosenzo

DISCUSSION PAPER SERIES

Social Proximity and the Erosion of Norm Compliance

Cristina Bicchieri
University of Pennsylvania

Eugen Dimant
University of Pennsylvania and CESifo

Simon Gächter
University of Nottingham, CESifo and IZA

Daniele Nosenzo
Aarhus University

ABSTRACT

Social Proximity and the Erosion of Norm Compliance*

We study how individuals' compliance with norms of pro-social behavior is influenced by other actors' compliance in a novel, dynamic, and non-strategic experimental setting. We are particularly interested in the role that social proximity among peers plays in eroding or upholding norm compliance. Our results suggest that social proximity is crucial. In settings without known proximity, norm compliance erodes swiftly because participants only conform to observed norm violations of their peers while ignoring norm compliance. With known social proximity, participants conform to both types of observed behaviors, thus halting the erosion of norm compliance. Our findings stress the importance of the broader social context for norm compliance and show that, even in the absence of social sanctions, compliance can be sustained in repeated interactions, provided there is group identification, as is the case in many social encounters in natural and online environments.

JEL Classification:	C92, D64, D9
Keywords:	norm compliance, social norms, social proximity

Corresponding author:
Eugen Dimant
University of Pennsylvania
311 Claudia Cohen Hall
249 S 36th Street
Philadelphia, PA 19104
USA

E-mail: edimant@sas.upenn.edu

1. Introduction

Decades of experimental research have challenged the notion that individuals are exclusively motivated by material self-interest, showing instead that in many situations they are prepared to sacrifice their own material well-being to increase the well-being of others. One important reason why individuals behave pro-socially is that, in most societies, there are *social norms* that constrain our self-interest and promote other-regarding behavior.[1]

However, recent experimental evidence suggests that pro-social behavior, and the norms that support it, are fragile. In settings where individuals can observe the behavior of their peers, seeing examples of norm violations has large negative effects on the individual's willingness to comply with the norm (Keizer et al., 2008; Gino et al., 2009a). In contrast, seeing examples of norm compliance has weaker positive effects on compliance, particularly when norm-following runs counter to material self-interest (Thöni and Gächter, 2015; Charness et al., 2019). This asymmetric response to information about others' behavior implies that norm compliance and pro-social behavior tend to wane in settings where individuals can observe each other's behavior (Dimant, 2019; Gächter et al., 2020).

These results paint a bleak picture for societal outcomes that depend on norms of pro-sociality. The picture, however, is incomplete. Most of the evidence on the asymmetric effects of bad and good examples comes from settings where individuals interact anonymously in groups of strangers, without receiving any signal of the *social (dis)similarity* of the individual to the peers one interacts with. In natural environments, however, people can typically assess the extent to which they share common traits, characteristics, and identities with those they observe (subsequently referred to as "social proximity"). In this paper, we explore whether signals of social proximity can counteract the disproportionate influence of bad examples on norm-following and thus reduce the erosion of compliance.

One reason why social proximity may play a role in moderating the influence of bad examples is that recognizing similarities between self and others brings forth a process of group identification, whereby the individual feels that she and her peers belong to the

---

[1]See, e.g., Fisman and Miguel (2007); Andreoni and Bernheim (2009); Krupka and Weber (2013); Reuben and Riedl (2013); Schram and Charness (2015); Gangadharan et al. (2016); Kimbrough and Vostroknutov (2016); Krupka and Croson (2016); Gächter et al. (2017); Kocher et al. (2017); Krupka et al. (2017); Barr et al. (2018); Kimbrough and Vostroknutov (2018); Bicchieri and Dimant (2019); Chang et al. (2019); Bicchieri et al. (2020a,b); Bursztyn et al. (2020); d'Adda et al. (2020); Dimant (2020).

same social group and attaches emotional significance to group membership (Tajfel, 1982). When an individual identifies with the group, she will seek to undertake behaviors that are in line with the expectations of what constitutes "appropriate group behavior" and avoid behaviors that signal disloyalty or disagreement with the group (Hogg, 1992; Akerlof and Kranton, 2000; Perkins, 2002; Chen and Li, 2009). To the extent that the perception of what constitutes appropriate group behavior in a given situation is informed by how other group members behave in that situation, this process of group identification will promote *behavioral conformity* within the group. That is, when the individual identifies with the group, she may pay attention to *both* examples of compliance and deviance, even when following these examples is materially costly, because any deviation from the group standard signals lack of commitment to the group and hence may be disapproved by other group members. Without group identification instead, examples of deviance may dominate examples of compliance because deviating from prosocial norms is typically materially advantageous for the individual (Charness et al., 2019). Thus, through group identification, social proximity could offset the tendency for bad examples to outweigh the influence of good ones.

To test this conjecture under controlled conditions, we designed a laboratory experiment using a Take-or-Give (ToG) donation game (which borrows elements from List, 2007 and Bardsley, 2008) in which decision-makers can either give money to a charity, take money from it, or retain the initial equal allocation of endowments between self and charity. This novel game allows us to study, within a unified, non-strategic setting, both pro-social behavior (giving to the charity) and anti-social behavior (taking money from the charity). Moreover, using a norm-elicitation experiment, we can show that our experimental participants recognize giving money to the charity as a social norm in the ToG game, and taking money as a norm violation. Thus, the ToG game represents a perfect paradigm to investigate the fragility of norm-sustained pro-social behavior.

We ran a repeated version of the ToG game where subjects make independent donation decisions across 20 periods. Between treatments, we systematically varied whether subjects can or cannot observe whether other decision-makers have given or taken money from a charity in previous periods. In the baseline treatment, "NoObservation", players received no feedback about others' ToG decisions in any of the 20 periods of the game. Thus, in this treatment, it is likely that subjects' decisions are purely based on the initial social norm that prevails in the ToG game and their own willingness to comply with it, and that these decisions remain relatively stable across the 20 periods.

In our first treatment, "Observation", we informed subjects after each round about the ToG decisions of other subjects in the same session. By observing these decisions, subjects can learn whether others comply with the norm of giving and the extent of any violations. This may lead subjects to revise their initial donation behavior. While initial decisions may be exclusively based on a subject's perception of the norm and his or her own willingness to comply with it, subsequent decisions can be affected by the comparison between the subject's and his/her peers' decisions. Based on the evidence discussed earlier, we expect these comparisons to be disproportionately affected by examples of norm violations (which in the ToG game are materially advantageous since taking from the charity increases a subject's payoff), so that the initial level of pro-social behavior in the ToG game may inexorably decline across the 20 periods of the game.

In our final and most novel treatment, "ObservationSP", we combined observability of peers' decisions with social proximity. In this treatment, subjects can not only repeatedly observe the ToG decisions of their peers, but they also learn whether they are similar or dissimilar to these peers in an irrelevant (to the context) and minimal characteristic (knowledge of a sport team victory).[2] As discussed above, sharing a common characteristic, even if it is minimal, may trigger a process of group identification so that subjects may conform to peers' behavior even when doing so is materially costly. Thus, we conjecture that the process of erosion of norm compliance that we expect to observe in the Observation treatment may be substantially reduced in ObservationSP.

Our results confirm our conjectures. We find that in the NoObservation treatment, donation rates to the charity are stable over time. This is an important benchmark result because it shows that increasing experience with the ToG donation game does not *per se* erode norm compliance. By stark contrast, donations in Observation decline by about 20% compared to donations in NoObservation. In line with previous findings in the literature, this decline is due to the asymmetric impact of observing compliance with (and violation of) the norm of giving. Individuals strongly reduce the amount they donate to charity when they observe that others take money from the charity, yet they do not increase the donated amount when they observe that others give to charity. Most importantly, knowledge of social proximity strongly moderates ToG donation behavior. In ObservationSP, the average donation is not significantly different from our NoObservation baseline. The reason is

---

[2]Note that our design implements the decisions of others in a non-strategic and anonymous way in all treatments, so that there is no monetary or strategic impact from learning others' behavior or characteristics.

that, in the presence of known social proximity, individuals respond to *both* observed norm compliance and deviance, thus stabilizing donations roughly at their initial level.[3]

These results show that observation of peer behavior, and the social proximity to those peers, have a strong influence on the dynamics of norm compliance. As in previous studies, and in our new ToG game, we find that observation of peer behavior leads to an erosion of norm compliance: seeing examples of norm violation induces individuals to lower compliance, whereas seeing examples of norm compliance does not lead to an analogous upward revision of compliance. An explanation for this asymmetric effect is that in our setting, examples of norm violations bolster selfish motives whereas examples of compliance require the individual to overcome their narrow self-interest.

Importantly, however, we show that the erosion of norm compliance strongly depends on the degree of known social proximity between individuals. When people are aware of social similarities with their peers, we observe strong responses not only to norm violations, but also to examples of norm compliance. This mitigates norm erosion in comparison to groups where individuals cannot assess their social similarity with other group members. This result highlights the importance of the broader social context in driving norm compliance, and shows that studying norm compliance (and pro-social behavior, more generally) only in anonymous and socially sterile settings overestimates the erosion of compliance. It suggests that punishment of norm violations, frequently seen as crucial to sustain norms (Coleman, 1994; Sutter et al., 2010; Fehr and Schurtenberger, 2018; Dimant and Gesche, 2020), is not always needed: even minimal social proximity can prevent the erosion of norm compliance.

Our paper proceeds as follows: Section 2 details the experimental design and hypotheses, Section 3 illustrates the results, and Section 4 concludes.

2. Experiments

We present data from a total of $N = 1,590$ participants across four distinct experiments (all between-subject designs in that no subject participated in more than one experiment):

1. *Behavioral experiment* utilizing the ToG donation game: $n = 842$ collected in the laboratory at the University of Pennsylvania.

---

[3]Interestingly, we also show that there is heterogeneity in the extent to which subjects respond to feedback about others' behavior. The biggest changes in behavior come from initial norm-compliant subjects who give to charity. In contrast, initial takers are relatively insensitive to feedback about others' behavior. See Figure A.6 in the Appendix.

2. *Norm elicitation experiment* (following the procedure by Bicchieri and Chavez, 2010) to establish which norm exists in our context: $n = 110$ collected in the laboratory at the University of Pennsylvania.

3. *Two additional norm elicitation experiments* establishing the robustness of the norm identified in Experiment 2 (the first elicitation followed the same procedure as in Experiment 2; the second elicitation followed the procedure by Krupka and Weber, 2013): $n = 464$ collected from a sample of the general population recruited online.

4. *Normative expectations change experiment* to better understand the mechanisms of changes in norm compliance (following the procedure by Bicchieri and Chavez, 2010): $n = 174$ collected in the laboratory at the University of Pennsylvania.

For the purpose of exposition, in the main text we provide a detailed discussion and illustration only of the results of Experiment 1 (ToG donation game). Accordingly, we limit our discussion to the key insights from Experiments 2, 3, and 4 in the main paper and provide the details of those experiments in the Online Appendix.

2.1. Behavioral Experiment: Design and Data Collection

In the ToG game, each subject makes a donation decision towards a charity. There are three possible charities available and the subject chooses one of the following to be paired with: Doctors Without Borders, World Wildlife Fund, or UNICEF. At the start of the game, the subject and the charity are both provisionally endowed with 100 ECUs (with 10 ECUs = \$1). The action space of subject i is represented by $x_i \in [-100, 100]$, where x_i is an integer representing the monetary amount which the ith subject decides to take from or give to the charity. The value 0 indicates no change to the initial equal split in endowments. The game is thus a variant of the dictator game that includes take options (e.g., List, 2007; Bardsley, 2008; Korenok et al., 2014), but where the recipient is a charity instead of another subject (as in Eckel and Grossman, 1996; Exley, 2015; Grossman and Eckel, 2015; Bolton et al., 2019; Dimant, 2019).[4]

This game allows us to study, in a unified framework, both pro-social behavior (giving to the charity) and anti-social behavior (taking from the charity). We conjectured

---

[4]Dimant (2019) studies a *one-shot* static version of the ToG game. As we explain below, in our paper we instead study a *multi-period* version of the game. The crucial advantage of our approach is that we can observe how norm compliance adapts dynamically to information about peer behavior. This allows us to study the factors that may lead to the erosion of norm compliance in settings where people can interact repeatedly and observe each other's behavior, as they do in most natural and online environments.

that behavior in this game is likely to respond to a *social norm of charitable giving* that prescribes that donating to the charity is socially appropriate and taking from charity is socially inappropriate. This is important because the focus of our study is on the erosion of *norm-driven* pro-social behavior. That is, the key aim of our paper is to study how observing other people's behavior affects the individual's decisions to comply with norms of pro-social behavior. To empirically confirm that behavior in the ToG game is indeed norm-driven, we conducted a series of norm-elicitation experiments measuring the extent to which subjects perceive giving to charity as socially appropriate and taking from the charity as socially inappropriate. These experiments are described in the Online Appendix. We find that participants recognize that giving to charity is the right thing to do in the ToG game (see Figures A.7; A.9; A.10) in all variants of our norm-elicitation experiments (with university students as well as members of the general population recruited online, using the Krupka and Weber, 2013 elicitation procedure as well as using the alternative procedure proposed by Bicchieri and Chavez, 2010). Moreover, among university students, there is the additional expectation that the norm will be followed in the sample we used in the behavioral experiment. Thus, in the following, we will refer to giving to the charity as *norm-compliant* behavior and taking from the charity as *norm-violating* behavior.

To study the erosion of norm compliance, in our behavioral experiment we recruited $N = 842$ participants (University of Pennsylvania students) to make choices in the ToG donation game. Participants were on average 22 years old and 70 percent were female. The experiment was conducted in zTree (Fischbacher, 2007) and ran in 52 sessions in the Behavioral Ethics Lab at the University of Pennsylvania across three between-subject treatments, described in more detail below.

In each treatment, every participant went sequentially through the same four parts of the experiment (plus a payment phase) that are illustrated in Figure 1. Participants were aware that the experiment consisted of multiple parts, but the details of each new part were only revealed upon completion of the previous parts. All decisions in the experiment were collected anonymously, i.e., none of the subjects' decisions could be linked to the respective individual (see Online Appendix for instructions).

Part I – Social Proximity Question
At the beginning of the experiment, all participants were asked one incentivized knowledge question (the year in which Philadelphia's baseball team won its last World Series). Participants were paid \$1 at the end of the experiment if they provided the correct answer (2008)

| PART I: Pre-Experiment | Incentivized social proximity question (what year the Phillies won the last World Series) |

| PART II: 1 Period of Individual Decision | PART III: 19 Periods of Individual Decisions Within a Group |

PART II: 1 Period of Individual Decision

One of three actions:

Take money away from charity for
personal gain
(Norm Violator)

Or

Do nothing and retain status quo
(Abstainer)

Or

Give money to charity at own expense
(Norm Follower)

PART III: 19 Periods of Individual Decisions Within a Group

Participants are randomly assigned to group of size 2 or 4 and continue to make individual decisions towards their charity across all 19 periods. Payoffs are independent between group members and periods.

Baseline
"NoObservation"
(N = 176) — ? — Group members do not observe other's ToG decisions

Treatment 1
"Observation"
(N = 268) — Observation — Group members observe other's ToG decisions in every period

Treatment 2
"ObservationSP"
(N = 398) — Observation + Social Proximity — Group members observe answer to social proximity question + other's ToG decisions in every period

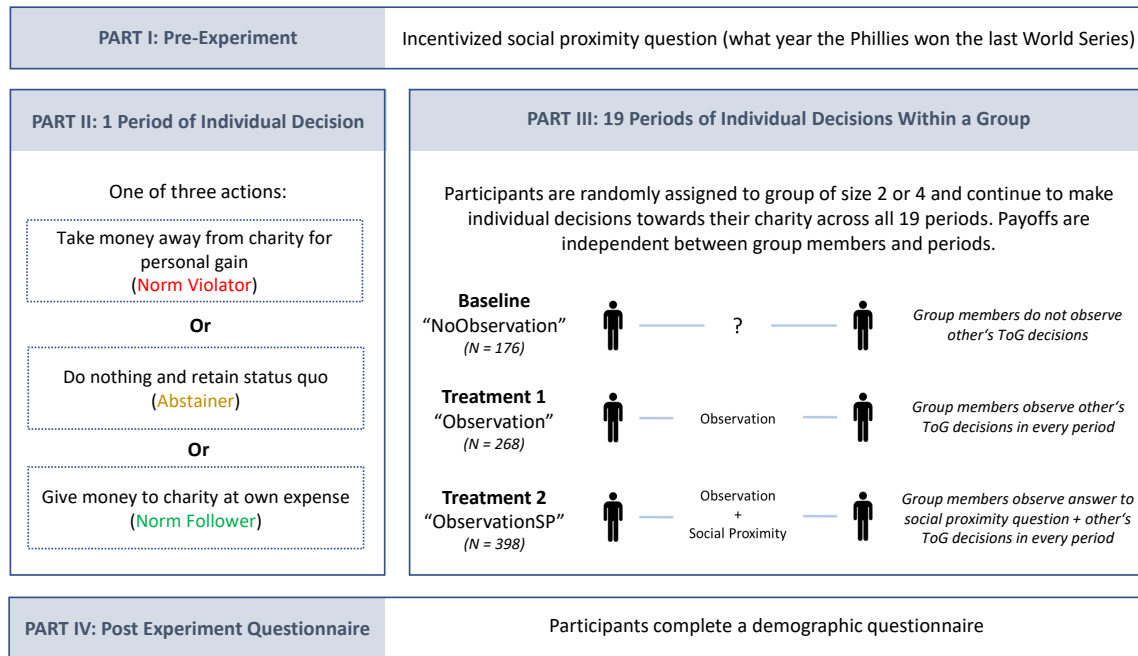| PART IV: Post Experiment Questionnaire | Participants complete a demographic questionnaire |

Figure 1: Experimental Design.

and \$0 otherwise. Overall, about 27% of our participants knew the correct answer. In the ObservationSP treatment (for details, see below), the answers to the knowledge question were used to give subjects a signal of their *social proximity* with other participants in the experiment on one specific dimension of their identity (being a supporter of Philadelphia's baseball team). At the end of Part II, participants were randomly assigned to groups and informed about how each member of their group had answered the question (and thus knew the extent to which they could be considered supporters of the Philadelphia team).[5] In the other treatments, participants were also asked the knowledge question, but were not

---

[5]To ensure that our social proximity signal is sufficiently strong, we decided to use a knowledge question about sports rather than an arbitrary grouping following the minimal-group paradigm literature (Tajfel and Turner, 1979). It has been shown that shared positive collective experiences (e.g., sports events) drive group identification in unrelated strategic contexts (Depetris-Chauvin et al., 2020). We tested (n=300 Amazon Mechanical Turk participants) the validity of our social proximity choice by comparing it to a standard minimal group marker using the 'Inclusion of the Other in the Self' scale (Aron et al., 1992; Gächter et al., 2015). After expressing their preference for sports or paintings, participants were asked to imagine being matched with someone who had the same sports/paintings preference as them and rate their closeness on the IoS scale from 1-7. Result: the sports identity achieved a significantly higher social identification than the minimal group choice (p<0.01, see Figure A.1 in the Online Appendix for details).

informed about other group member's answers until the very end of the experiment. In the ObservationSP treatment, this information was instead given between Parts II and III.

Part II - Individual Decision (one period)

Next, all participants made one ToG decision. Participants first selected their preferred charity from a list and then chose how much to take or give (from 0 ECUs up to +/-100 ECUs) from/to this charity. It was public knowledge that each individual's decision did not affect other participants' payoffs or the payoff of a charity besides their own. This design choice removes any payoff interdependency between participants. This is a key advantage of the ToG game over other games that have been used to study how behavior is affected by the observation of peers' actions (e.g., public good games; ultimatum games), as it allows to study the effects of information about peer behavior in a setting where there are no strategic reasons to respond to such information.

Part III - Individual Decisions (nineteen periods)

Following the individual decision in Part II, for periods 2 to 20, participants were randomly assigned to a group that consisted of either two ($N = 354$) or four ($N = 488$) participants. The group allocation remained constant throughout the experiment. We varied group size because existing evidence suggests that group size affects the pressure to conform with group norms (see Bond, 2005). Each participant continued to make the same type of individual decision as in Part II towards his or her charity for another 19 periods. In each period, both the subject and the charity were endowed with 100 ECUs each to avoid any path-dependent carryover effects from previous periods. Our treatments varied as follows.

In our baseline condition (NoObservation), participants were placed in groups of two or four members, but no behavior was observable by or revealed to any other participant. We study the natural evolution of compliance with the norm of giving across the 20 periods of Part II and III of the experiment.

Our first treatment (Observation) is identical to the baseline condition except that in each period of Part III, participants received information about the ToG decisions of the other group members in the previous periods of the experiment, including their decisions in Part II of the experiment. This feedback information was presented in the form of a history table to allow participants to track the ToG decisions of each participant in their group over the whole duration of the experiment. At no point during the experiment were participants able to observe the ToG decisions of another participant who was not in their

group. The comparison between Observation and NoObservation allows us to assess the effect that repeated exposure to information about others' behavior has on compliance with the giving norm in the ToG game. Finally, our second treatment (ObservationSP) is identical to Observation, except that in addition to observing group members' ToG decisions, participants also received information about their social proximity to their group members through the answers to the knowledge question of Part I. This information was displayed right below each of the group members' history tables in the form of an "answered the question correctly/incorrectly" indicator. We use this treatment to assess how the ToG donation decisions are affected by the social proximity of other group members.

Part IV - Post-Experimental Questionnaire
We elicited general participant information, such as age and gender, alongside individual behavioral attributes, such as risk attitudes and opinions about charitable organizations, all of which we control for in our regression analyses.

Payment
All participants were paid a show-up fee of $10 in addition to their earnings from decisions made in the experiment. At the end of the experiment, we randomly selected one subject in each session and implemented exactly one (randomly-chosen) decision of this subject from the 20 decisions he/she had made during Part II and III of the experiment. We used this randomly chosen decision to compute the subject's own payoff as well as the charity's payoff. The randomly selected participant also determined the charity which the money was transferred to/from. The decision of every other subject in the same session did not count towards his/her own and the charity's payoff. Instead, each of these subjects received exactly 100 ECUs regardless of their allocation decision towards their charity in any of the periods. Participants were informed of this payment procedure at the beginning of the experiment. We chose this type of 'pay one' payoff structure to further eliminate from the design any type of interdependencies between subjects, not just in terms of their material payoffs, but also in terms of potential social payoffs; e.g., concerns about the charity or strategic substitution effects ('I can take because you give and balance out the payoffs'). This setup is in line with suggestions made by Charness et al. (2016) and retains incentive compatibility as theoretically argued by Azrieli et al. (2018). The experiment lasted about 45 minutes, yielding an average hourly pay of $18.

2.2. Behavioral Experiment: Hypotheses

The main aim of our paper is to study how compliance with the norm of giving to charity varies across the 19 periods of Part III of the behavioral experiment. In particular, we investigate how observing group members' decisions to give/take to/from the charity affects an individual's willingness to comply with the norm of giving to charity in the two treatments with and without signals of social proximity (Observation and ObservationSP).

Following the literature on peer effects in pro-social behavior (e.g,, Gino et al., 2009a; Charness et al., 2019; Gächter et al., 2020), when individuals interact anonymously and in the absence of signals of social proximity, we expect an asymmetric effect of information about peer behavior on norm compliance: evidence of norm-violating ToG decisions will have a stronger impact than evidence of norm compliance. Thus, we expect a decline in compliance with the giving norm across the 19 periods of Part III in Observation, compared to NoObservation where peer information is not available. Our first hypothesis is as follows:

H_1: *In Observation, individuals react more strongly to examples of norm violation (taking) than examples of norm compliance (giving). Over time, this leads to a decline of compliance with the norm of giving compared to NoObservation.*

We conjecture that this asymmetric effect of peer information (and the subsequent decline in norm compliance) is due to a self-serving interpretation of the information available to the individual, as they may ignore evidence of norm compliance because conforming with it is materially costly. However, we expect social proximity to reduce this self-serving use of information about peer behavior: ignoring evidence of norm compliance may be costly when this evidence comes from the behavior of socially proximate group members, since non-conformity may be negatively judged by the group (Hogg, 1992). Thus, we expect the decline of norm compliance to be reduced (or completely halted) among groups where individuals can observe signals of social proximity with other group members. This conjecture implies a smaller decline in norm compliance in ObservationSP compared to Observation. Moreover, we expect that this effect is driven by the most homogeneously proximate groups in ObservationSP, i.e., the groups where all members have given the same answer to the social proximity question of Part I of the experiment.

H_2: *Social proximity reduces the asymmetric effect of peer information: subjects in ObservationSP will respond to examples of both norm compliance and violation, thus reducing the erosion of norm compliance. This effect will be driven by subjects in ObservationSP that belong to groups that are most homogeneously socially proximate.*

3. Results

3.1. Behavioral Experiment

3.1.1. Behavior in Part II - Initial Compliance With the Norm of Giving

The first period of the ToG game (Part II of the experiment) is the same in all treatments: subjects make a ToG decision without receiving any information about the behavior of others or their social proximity. In the Observation and ObservationSP treatments, participants did not know that their decisions in Part II would be displayed to other group members in Part III. We can use these initial donation decisions to establish the extent to which the norm of giving that exists in the ToG game (as established in our norm-elicitation experiments; see Online Appendix) is followed when subjects have to rely on their original expectations about the appropriateness of compliance.

Table 1 displays the average ToG donation in Part II of the experiment across the three treatments. Due to lack of group size differences, we pooled the data across group sizes (see Table 2). The table also displays the percentage of participants in each treatment who gave, took, or abstained from changing initial endowments. We label participants who gave to charity "norm followers", those who took money from charity "norm violators", and those who neither gave nor took as "abstainers".

	Average amount given/taken	Percentage of initial takers (norm violators)	Percentage of initial equal split (abstainers)	Percentage of initial givers (norm followers)
NoObservation	-18.9 (57.1)	45%	36%	19%
Observation	-25.2 (55.9)	44%	40%	16%
ObservationSP	-21.8 (51.1)	42%	41%	19%

Table 1: ToG decisions in Part II of the experiment. Standard deviations in parentheses.

Participants on average took between 18.9 and 25.2 ECUs from the charity. In all treatments, the majority of participants took from the charity (frequencies range between 42% and 45%, depending on the treatment). Only a minority of participants gave to the charity (between 16% and 19% depending on the treatment). As expected per proper treatment randomization, we find that behavior in Part II does not differ across treatments (Kruskal-Wallis test, df $= 2$, p $= 0.703$). Our data shows that there is a significant gap between subjects' initial behavior and the norm measured in our norm-elicitation experiment. The

latter clearly indicates that a norm of giving exists in the game, but the behavior in the experiment shows that compliance with the norm is quite low. This suggests that, although subjects recognize giving to charity as a socially appropriate behavior in our setting, the costs of compliance outweigh the material benefits of violations.

3.1.2. Behavior in Part III - The Dynamics of Norm Compliance

We now explore how compliance is affected when subjects are given information about the behavior of others in the game, which allows them to revise their expectations about others' compliance towards the actual behavior in the game. Figure 2 (left panel) shows the evolution of average donations over the 19 periods of Part III across our treatments. Figure 2 (right panel) shows ToG donations averaged across all 19 periods. In both cases, the figures reflect average donations normalized relative to behavior in Part II (which is displayed as period 0 in Figure 2, left panel). A positive value indicates that, relative to the average in Part II, the account balance of the charity has increased, and a negative value indicates that the account balance of the charity has decreased.
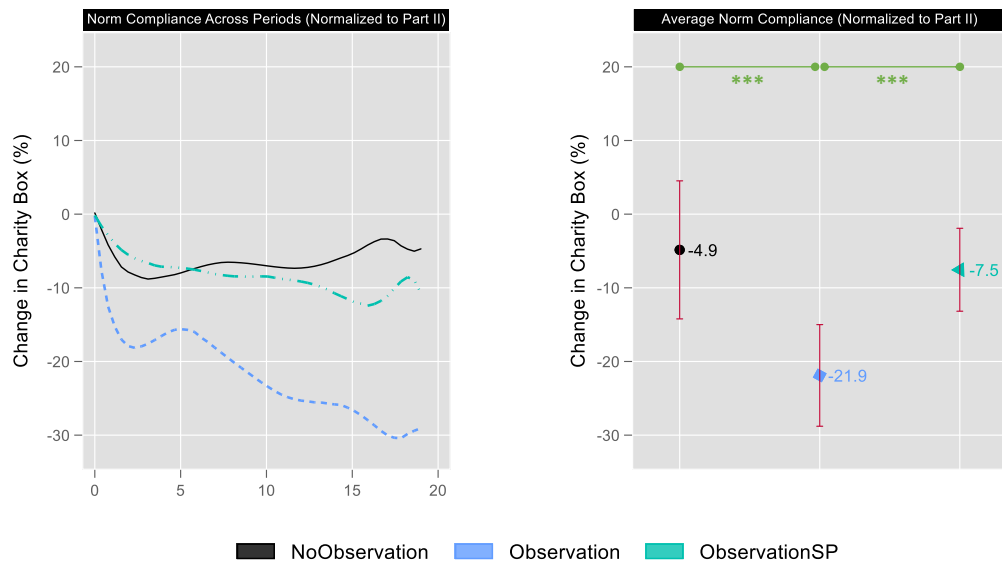

Figure 2: Left panel: Subject's behavior across treatments and periods in Part III, normalized to the average behavior in Part II (period 0 in the graph). For purposes of readability, lines represent kernel-weighted local polynomial smoothing of degree 5. Right panel: Average subject's donations across treatments, averaged over the 19 periods of Part III. Whiskers indicate 95% confidence intervals. Stars indicate significant differences using two-sided Mann-Whitney U-tests and group averages as unit of observation. P-values adjusted for multiple hypothesis testing (Benjamini and Hochberg, 1995). *p<0.1, **p<0.05, and ***p<0.01.

In NoObservation, where participants did not receive any additional information about the behavior of other group members, donations in Part III are similar to donations in Part II. On average, across the 19 periods of Part III, the balance of the charity was reduced by an additional 4.9 ECUs relative to Part II. Using two-sided Wilcoxon sign-rank tests with group averages as observations, we cannot reject the null hypothesis that average donations in Part III of NoObservation are identical to donations in Part II ($z = -0.921$, $p = 0.357$). In the absence of new empirical information, we do not observe a significant change in compliance with the norm of giving across the 19 periods of Part III.

In contrast, when participants received information about other group members' behavior, the average amount donated to the charity dropped relative to donations measured in Part II. Using two-sided Wilcoxon sign-rank tests with group averages as observations, we detect significant differences between Part II and Part III donations for Observation ($z = -5.487$, $p < 0.001$) and ObservationSP ($z = -2.370$, $p = 0.018$).

However, the decline in compliance appears to be much stronger in Observation, where participants took on average an additional 21.9 ECUs relative to Part II, compared to ObservationSP, where the balance of the charity was reduced by an additional 7.5 ECUs. A Kruskal-Wallis test confirms that there are significant differences between the three treatments in Part III of the experiment ($df = 2$, $p < 0.001$). We perform bilateral treatment comparisons using two-sided Mann-Whitney U-tests and group averages as a unit of observation, with p-values adjusted for multiple hypotheses testing (Benjamini and Hochberg, 1995). Across all periods of Part III, we find that the average donations are significantly different between Observation and NoObservation ($z = -3.732$, $p < 0.001$) and also between Observation and ObservationSP ($z = 3.524$, $p < 0.001$). No significant difference is detected between NoObservation and ObservationSP ($z = 0.877$, $p = 0.381$).[6]

We test the robustness of our findings by performing multivariate regression analyses that include controls for observable differences across treatment groups. In Table 2, column 1 reports the results of random-effects GLS regressions where the dependent variable is the amount given to the charity by subject i in period t of Part III of the experiment. The independent variables are treatment dummies (Observation and ObservationSP, using NoOb-

---

[6]Importantly, our main result is not driven by the disproportionate evidence of taking-behavior in Part II. In fact, as illustrated in Figure A.2 in the Online Appendix, we observe a substantial decay in norm compliance even for those groups in which such behavior was initially in the minority. This additional robustness check emphasizes the generalizability of the observed dynamics.

servation as base category), a dummy for group size (using group size 2 as base category), a variable measuring the amount contributed by the participant in Part II (to control for a participant's underlying predisposition to give or take from the charity), a period variable, and various other controls (including the correct answer to the social proximity knowledge question and socio-demographic controls collected in the post-experimental questionnaire).

The regression in column 1 of Table 2 corroborates the results reported above. Relative to the NoObservation treatment, the account balance of the charity was reduced significantly in the Observation treatment; the difference is significant at the 1% level. Moreover, a test of equality of coefficients between Observation and ObservationSP confirms that the account balance of the charity was reduced significantly in Observation ($\tilde{\chi}^2(1) = 19.30$, p < 0.001). However, we do not observe any significant difference between NoObservation and ObservationSP. We also do not detect any difference in behavior in Part III of the experiment across participants who were in groups of size 2 or 4. We do not observe any general effect of group size on taking behavior; a separate regression in which we interact group size with treatments does not yield any significant differences either.

Our results so far show that when participants only receive feedback about the behavior of other group members, norm violations spread, in line with our first hypothesis H_1 about the effects of information about others' behavior. In contrast, when participants receive both feedback about the behavior of other group members and information about their social proximity, the decline in compliance is substantially reduced. Behavior is at a level comparable to NoObservation, which supports our hypothesis H_2.

Why does norm compliance not significantly decline when information about social proximity is available? At the heart of H_1 and H_2 is the idea that, in the absence of knowledge of social proximity, participants mainly respond to examples of norm violations. Instead, when group members have knowledge of social proximity, participants also respond to examples of norm compliance in addition to norm violation, leading to a zero net effect.

To better understand the drivers of the decline in norm compliance across our treatments, in column 2 of Table 2 we augment the random-effects GLS regression of column 1 with variables capturing the different types of information that participants were exposed to in the 19 periods of Part III, as well as interactions between these variables and the treatment dummies. For the purpose of our analysis, we follow Fischbacher and Gächter (2010) and introduce variables lagged by one period that capture the effect of information

DV: Amount Change in Charity Box	(1)	(2)
Treatment		
(Base level: NoObservation)		
Observation	-18.02***	-4.26
	(3.51)	(3.95)
ObservationSP	-3.67	0.60
	(3.09)	(3.15)
Groupsize = 4	-2.81	-1.73
	(2.64)	(2.36)
Initial Individual Donation (Part II)	0.67***	0.67***
	(0.02)	(0.02)
Round	-0.35***	-0.25***
	(0.08)	(0.07)
Neg. Avg Don. t-1		0.05
		(0.03)
Neg. Avg Don. t-1 * Observation		-0.29***
		(0.05)
Neg. Avg Don. t-1 * ObservationSP		-0.17***
		(0.05)
Pos. Avg Don. t-1		-0.01
		(0.04)
Pos. Avg Don. t-1 * Observation		-0.02
		(0.10)
Pos. Avg Don. t-1 * ObservationSP		0.21***
		(0.07)
Zero Avg Don. t-1		-1.18
		(1.45)
Zero Avg Don. t-1 * Observation		0.69
		(3.20)
Zero Avg Don. t-1 * ObservationSP		2.60
		(2.18)
Constant	-36.95***	-39.85***
	(7.20)	(7.22)
Controls	Yes	Yes
N.	15998	15998
N. Clusters	299	299
R^2	0.47	0.49

Table 2: Random-effects GLS regressions. Dependent variable is donation of subject i in period t of Part III of the experiment. Controls are: gender, age, whether or not the proximity knowledge question has been answered correctly, a measure of self-control taken from Tangney et al. (2004), a variable measuring the (self-reported) liking for charities, a measure for the ability to anticipate future consequences of current behavior taken from Strathman et al. (1994), and a self-reported measure of risk preferences taken from Dohmen et al. (2011). We always use the same set of controls in all of our regressions. Because our analysis does not focus on the role of these controls and in order to allow for better readability, the estimates of the controls are not reported separately in Table 2, but are available upon request. Robust standard errors reported in parentheses (standard errors clustered at the group level). Stars indicate significant differences at the conventional levels of *p<0.1, **p<0.05, and ***p<0.01.

diffusion within groups.[7] We distinguish between three types of information. First, we use the variable "*Neg. Avg Don. t-1*" to capture the effect of receiving feedback that other group members on average took money from the charity in the previous period (i.e., participants observed, on average, an example of norm violation). Thus, this variable is equal to the absolute value of the (lagged) average amount transferred to the charity by other group members if this amount is strictly negative, and it is equal to 0 otherwise. Second, we use the variable "*Pos. Avg Don. t-1*" to capture the effect of receiving feedback that the other group members on average gave money to the charity in the previous period (i.e., an example of norm compliance). This variable is equal to the (lagged) average amount transferred to the charity by other group members if this amount is strictly positive, and 0 otherwise. Finally, to capture the effect of observing an average zero transfer to the charity in the previous period, we use the dummy variable "*Zero Avg Don. t-1*", which takes value 1 if the amount transferred was zero, and 0 otherwise. We then interact each of these variables with our treatment dummies.

The interactions between the lagged donation variables and the Observation dummy capture the differential impact of information about peer behavior in Observation relative to NoObservation.[8] The estimates reveal that observing norm-compliant behavior, or receiving feedback that on average other group members left the charity's endowment untouched, does not statistically significantly affect a participant's decision to give or take money to/from the charity.[9] In contrast, observing norm-violating behavior has a strong and negative effect on donations. Our estimate suggests that receiving feedback about other group members taking 1 ECU from the charity reduces the amount that is donated on average by -0.29 + 0.05 = -0.24 ECUs (see Table 2, column 2), which is significant at the 1% level.[10] An F-test shows that the effects of examples of violation are statistically significantly larger than the effects of examples of compliance ($\tilde{\chi}^2(1) = 18.41$, p < 0.001).

---

[7]These results are robust to also controlling for higher lags of t_{-2} and t_{-3} (available upon request).

[8]Given our interactions, the main coefficients of the lagged donation variables (*Neg. Avg Don. t-1, Pos. Avg Don. t-1*, and *Zero Avg Don. t-1*) measure the effects of the average amount transferred by other group members in NoObservation. This is a placebo test for the effects of peer behavior information since participants in NoObservation did not actually receive feedback about the amounts transferred by other group members during the experiment. As expected, the estimates show that in NoObservation information about others' behavior did not have any significant effect on donations.

[9]We cannot reject that *Pos. Avg Don. t-1 + Pos. Avg Don. t-1 * Observation = 0* (p = 0.870) or that *Zero Avg Don. t-1 + Zero Avg Don. t-1 * Observation = 0* (p = 0.810).

[10]We reject: *Neg. Avg Don. t-1 + Neg. Avg Don. t-1 * Observation = 0* ($\tilde{\chi}^2(1) = 34.40$, p < 0.001).

The interactions between the lagged donation variables and the ObservationSP dummy capture the impact of feedback about others' behavior when participants could also observe a signal of social proximity to the other group members. As in Observation, feedback that on average other group members abstained from changing the charity's endowment does not affect a participant's donation.[11] Also, as in Observation, observing norm violations has a negative impact on donations, with an estimated reduction of donations of -0.17 + 0.05 = -0.12 ECUs for each ECU that others took from the charity (see Table 2, column 2). This effect is significant at the 1% level.[12] However, in contrast to the Observation treatment, participants in ObservationSP also responded to examples of norm compliance. Receiving feedback that other group members give on average 1 ECU to the charity increases the amount donated by 0.21 - 0.01 = 0.20 ECUs, which is also significant at the 1% level.[13] An F-test does not reject the null that the effects of observed compliance and violation are of similar magnitude ($\tilde{\chi}^2(1) = 0.21$, p = 0.649).

Lastly, we observe important heterogeneity when applying a difference-in-difference analysis in the context of either observing norm compliance or norm violation. For norm violation, the interactions (-0.29 and -0.17) are significantly different from each other ($\tilde{\chi}^2(1) = 4.09$, p=0.043), indicating that the negative impact resulting from observing norm violations from socially proximate peers (ObservationSP) is smaller compared to observing peers in the absence of known social proximity (Observation). For norm compliance, the interactions (-0.02 and 0.21) are also significantly different from each other ($\tilde{\chi}^2(1) = 6.92$, p<0.01), indicating that the positive impact resulting from observing norm compliance from socially proximate peers (ObservationSP) is larger compared to observing peers in the absence of known social proximity (Observation). Thus, in both cases, knowledge of social proximity moderates the effect of observing peer behavior.

This analysis reveals that the fundamental difference between the treatments with observation lies in the fact that without information about social proximity, participants respond significantly only to examples of norm violation by reducing their donated amount. However, being informed about the social proximity of the other group members induces responses to observations of norm compliance as well as norm violation. Participants in the ObservationSP treatment reduce the amount donated when they observe others taking

---

[11]We cannot reject that *Zero Avg Don. t-1 + Zero Avg Don. t-1 * ObservationSP = 0* (p = 0.225).

[12]We reject: *Neg. Avg Don. t-1 + Neg. Avg Don. t-1 * ObservationSP = 0* ($\tilde{\chi}^2(1) = 13.23$, p < 0.001).

[13]We reject: *Pos. Avg Don. t-1 + Pos. Avg Don. t-1 * ObservationSP = 0* ($\tilde{\chi}^2(1) = 8.02$, p = 0.005).

from the charity, but they also increase the amount donated when they observe that others give money to the charity. The effects of observing others taking and giving to charity are roughly similar in magnitude. As a consequence, the net effect of others' behaviors on donations in this treatment is not significantly different from zero.

Overall, these results suggest that knowledge of social proximity can moderate the decline in norm compliance that would otherwise occur in interactions among anonymous strangers. In line with H_2, an implication of this finding is that the moderating effects of social proximity may be stronger among groups in which all members are similar in the relevant social dimension. We examine this conjecture in the regressions of Table 3 for ObservationSP in more detail. The regressions report the effects of observation of others' behavior (disaggregated in the same way as in Table 2) on donations in ObservationSP. In column 1, we only use observations from groups where either all group members had answered the knowledge question of Part I correctly (thus revealing that they were all likely to be fans of Philadelphia's baseball team), or where all group members answered it incorrectly (revealing that none of them were likely to be fans of Philadelphia's baseball team). In column 2, we instead use observations from groups where some group members had answered the question correctly, whereas others had answered it incorrectly. Hence, we expect social proximity to be higher among the former group than the latter. Based on this and our conjecture about the relation between proximity and observation, we expect the effect of proximity to be stronger among the more socially proximate groups.

The regression results confirm our conjecture. Among participants of socially proximate groups (column 1), we observe that donations are affected by examples of both norm compliance and violations. In contrast, in non-proximate groups (column 2), participants responded to examples of norm violations but did not significantly respond to examples of norm compliance, just as in Observation. Moreover, although not significantly different, the point estimates indicate the magnitude of the effects of violation and compliance are larger among proximate versus non-proximate groups (0.16 vs. 0.09 for examples of violations; 0.21 vs. 0.16 for examples of compliance). This result supports the conjecture that the mechanism behind the differential effects of information between Observation and ObservationSP is indeed related to the degree of known social proximity.[14]

---

[14]One could further distinguish between groups where all members gave a correct answer and groups where all members gave an incorrect answer. However, we have only a very small number of groups where all members gave a correct answer to the knowledge question. Because we do not have sufficient statistical power to draw reliable conclusions, we merely present an exploratory graphical representation in the Online

DV: Amount Change in Charity Box (only in ObservationSP condition)	(1) Proximate Groups	(2) Non-Proximate Groups
Negative Avg. Donation $_{t-1}$	-0.16***	-0.09***
	(0.06)	(0.03)
Positive Avg. Donation $_{t-1}$	0.21***	0.16
	(0.07)	(0.11)
Zero Avg. Donation $_{t-1}$	-2.24	3.83
	(2.35)	(2.44)
Groupsize = 4	-4.52**	-1.19
	(2.05)	(2.67)
Initial Individual Donation (Part II)	0.66***	0.65***
	(0.05)	(0.06)
Period	-0.16	-0.39***
	(0.14)	(0.14)
Constant	-15.43	-58.46***
	(15.98)	(16.94)
Controls	Yes	Yes
N.	3648	3914
N. Clusters	78	70
R^2	0.55	0.37

Table 3: Random-effects GLS regressions. Dependent variable is donation of subject i in period t of Part III of the experiment of ObservationSP since proximity was observable only in this treatment. Robust standard errors reported in parentheses (standard errors clustered at the group level). Column 1 uses observations from groups where all group members answered the knowledge question of Part I in the same way (either correctly or incorrectly). Column II uses observations from groups where group members gave different answers to the knowledge question. Controls are the same as in Table 2; see note to Table 2. Stars indicate significant differences at the conventional levels of *p<0.1, **p<0.05, and ***p<0.01.

3.1.3. Heterogeneity in the Erosion of Norm Compliance

An additional question that we address in this subsection is whether initial dispositions toward norm compliance mediate responses to observation of others' behavior. From a policy perspective, this analysis is interesting because it not only helps to understand who is more susceptible to the effects of observation, but to what extent their behavior changes and contributes to the erosion of norm compliance.

Based on the donation behavior in Part II of the experiment, we divided our participants into three groups: norm followers, violators, and abstainers. As previously shown in Table 1, between 16% – 19% of participants can be classified as norm followers, 42% — 45% as norm violators, and 36% — 41% as abstainers. To examine how each group of subjects is differently affected by observation, we consider their behavior across three

---

Appendix Figures A.3 and A.4. The suggestive evidence is in line with our discussion.

distinct phases of the experiment: Part II, where there was no information about others' behavior, and periods 1-10 and 11-19 of Part III, where, in the treatments with observability, they continuously received empirical information about others' ToG decisions in their group. Figure 3 plots separately for each type of subject and treatment, the distribution of donations made in these three phases of the experiment, depicted in blue, green, and red, respectively. The vertical dotted lines in each panel indicate the average donation behavior and the colors correspond to the respective phase (for more detail also Figure A.6 in the Online Appendix).
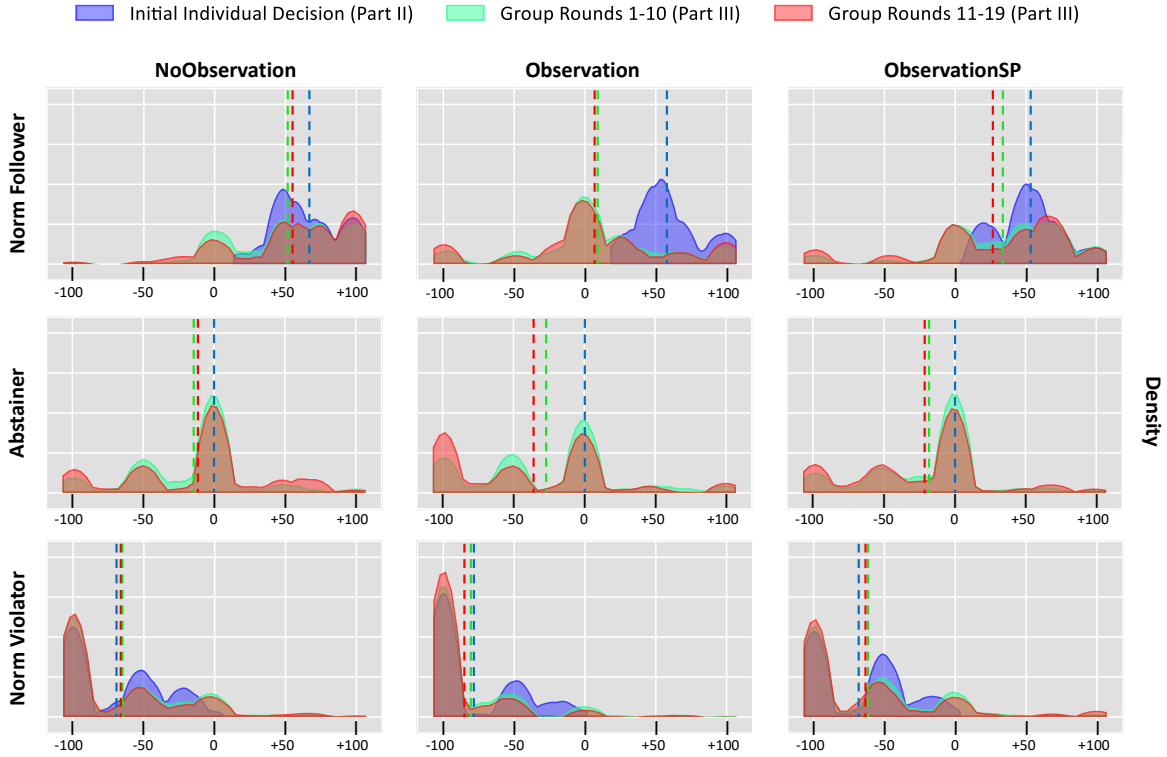

Figure 3: Shown are the distribution of donations (density) disaggregated by group of subject (norm violators, abstainers, norm followers, as defined by Part II initial behavior) and treatment. Vertical lines represent averages. Colors correspond to the respective phases. Blue: donations in Part II. Green: average donations across periods 1-10 of Part III. Red: average donations across periods 11-19 of Part III.

We have a number of interesting results. In NoObservation (leftmost column of Figure 3), most individuals only make small changes to their behavior between Part II and III of the experiment, which is expected given that participants are not observed or observing others. Norm followers give on average 65.3 ECUs to charity in Part II and 52.2 ECUs in Part III,

a significant decrease of 13.1 ECUs (MWU, p = 0.013). Abstainers, by definition, give 0 to the charity in Phase II of the experiment. In Phase III, their average donation drops to -13.0 ECUs, and the difference is statistically significant (MWU, p = 0.005). Finally, norm violators take on average 69.3 ECUs from charity in Part II, while the average amount taken across the 19 periods of Part III is 64.9 ECUs, which is not statistically significant (MWU, p = 0.140). The change in behavior of both norm followers and abstainers is larger than that of norm violators (both p <0.005); we find no difference between norm followers and abstainers (p = 0.994).[15]

In Observation (middle column of Figure 3), donations drop more dramatically. Again, we find the largest changes in behavior for norm followers and abstainers, while norm violators only minimally adjust their choices. Norm violators' donations drop from -78.9 ECUs in Part II to -82.7 ECUs in Part III (a statistically insignificant decrease of -3.8 ECUs, p = 0.220). Norm followers reduce their donations from +58.8 ECUs in Part II to +7.6 ECUs in Part III (a drop of -51.2 ECUs), and abstainers reduce their donations by -30.8 ECUs. Both effects are highly significant (p <0.001) and different from the change in behavior of violators (both p <0.001). We also find that the drop in donations of norm followers is significantly larger than that of abstainers (p = 0.011).

We find similar effects in ObservationSP (rightmost column of Figure 3), albeit smaller in magnitude. For norm followers, donations drop from +51.9 ECUs in Part II to +30.2 ECUs in Part III, a change of -21.7 ECUs that is highly significant (p < 0.001). Abstainers significantly reduce their donation by -20.1 ECUs (p < 0.001). For norm violators, donations go from -67.9 ECUs in Part II to -61.0 ECUs in Part III. This change of 6.9 ECUs is statistically significant (p = 0.009). Again, we find that the change in donations of followers and abstainers is larger than that of violators (p < 0.001), but we find no difference in donations between followers and abstainers (p = 0.764).

The take-away message from this analysis is that initial norm followers display the largest changes in behavior between Part II and III of the experiment, followed by initial

---

[15]To assess the statistical significance of the changes in behavior between Part II and III, we use OLS regressions, conducted separately for each treatment. The dependent variable measures, for each subject, the difference between the donation made in Part II and the average donation made in Part III (thus, the subject is the unit of observation in this analysis). We regress this on group dummies, and test whether, for each group, the estimated coefficient is significantly different from zero. The standard errors are clustered at the group level to account for potential interdependencies between subjects of the same group. We report two-sided p-values adjusted for multiple hypothesis testing (Benjamini and Hochberg, 1995).

abstainers, and then initial norm violators. The reason for this is straightforward: based on Figure 2, norm followers experienced the largest discrepancy between their own behavior and the behavior of their group members. In contrast, norm violators observed the smallest difference. This is in line with our opening discussion: we expect norm followers to be those who would have to make the largest (downwards) revisions to their expectations about others' norm compliance after receiving information about how other group members *actually* behaved. Since compliance with norms partly depends on whether others also comply (Bicchieri, 2006, 2016), their behavior is strongly influenced by what they observe. Indeed, in line with this argument, we also find that even among initial norm violators, the strongest behavioral adjustments to empirical information were made by those who were initially "moderate" in their taking behavior (took less than 50% of the charity endowment), as illustrated in Figure A.5 in the Online Appendix.

3.2. A Note on the Drivers of the Erosion of Norm Compliance

In this final subsection, we conclude with a discussion of our final experiment 4 in which we investigate the possible deeper mechanisms that can explain the observed erosion of norm compliance in our behavioral experiment. Here we only review the motivations and results of the last experiment (for a detailed discussion, see the Online Appendix).

As shown in our behavioral experiment, observing others' behavior induces individuals to change their own behavior. This suggests that an individual's willingness to comply with norms of pro-social behavior depends on whether other people also comply. Indeed, models of social norms (e.g., Bicchieri, 2006, 2016) posit that individuals have preferences for complying with norms that are conditional not only on whether others think that they ought to do so ("normative expectations"), but also on whether others do so themselves ("empirical expectations"). Thus, a first *direct* mechanism that can explain the results of our behavioral experiment is that observing others' behavior in the experiment induces participants to update their initial empirical expectations about others' compliance with the norm of giving, and hence their willingness to comply with it. As we have seen in the experiment, this process of belief updating may be biased towards self-interest in the absence of mechanisms — such as social proximity — that can nudge individuals to pay attention to examples of compliance as well as examples of violation.

However, observing others' behavior may also *indirectly* affect compliance by inducing participants to revise their beliefs about what constitutes socially appropriate behavior in the experiment. That is, when faced with examples of norm transgression, participants may

22

revise their normative expectations about what others think one ought to do in the game. This *indirect* effect of observation on normative expectations has been documented in other settings, e.g., Bicchieri et al. (2020a) and Gächter et al. (2020). To test whether this also occurs in our setting, we conducted a follow-up experiment, where we elicited participants' personal normative beliefs as well as their beliefs about what others consider appropriate behavior in the ToG game *twice*: before and after being shown how participants in the behavioral experiment actually behaved. Overall, we find compelling evidence that the erosion of norm compliance is associated with a downward shift in normative expectations, but not in personal normative beliefs (Figure A.11). Combined with our findings from the behavioral experiment, these results suggest that observing others' behavior has both a direct and an indirect effect on norm compliance, as observation of peers' behavior leads subjects to update both their expectations about what others think one ought to do and about what others will in fact do. Since norm compliance is conditional on such expectations (Bicchieri, 2006, 2016), revisions in these beliefs trigger a change in behavior.[16]

4. Discussion and Conclusion

Social norms are a fundamental component of social and economic life. Therefore, it is important to study conditions under which norm compliance occurs. In this paper, we focused on how observing others' norm compliance or norm violation influences individual compliance. To investigate norm compliance, we designed a novel, non-strategic Take-or-Give (ToG) donation game where people could give to charity, take from it, or abstain from changing the initial allocation between the self and the charity. Using a series of norm-elicitation experiments, we established that most people think taking from the charity is socially inappropriate, whereas giving to the charity is appropriate. We then examined the effect of letting individuals observe each other's behavior in a repeated version of the ToG game. Our behavioral results reveal a notable asymmetry in norm compliance: observing other anonymous individuals violating the norm (taking from the charity) increased the likelihood that the observers transgress as well. Observing norm compliance (donating to the charity), however, did not increase donations to the charity. In sum, observing norm violations by anonymous people eroded norm compliance in a way that was not compensated by observing norm compliance.

---

[16]With that, our results are in line with the work of Mussweiler (2001, 2003) on the effects of observation and group similarity on behavior.

While this asymmetry in reactions paints a bleak picture about norm compliance when other anonymous people can be observed, in most real-world interactions people have clues about their social proximity to the people they observe. Assessing similarities with others may bring forth a mechanism of group identification whereby the individual self-categorizes as a member of a particular group and recognizes that his/her peers are also part of the same group. Group identification may nudge individuals to pay attention to observations of compliance as well as violation, and thus promote *symmetrical* behavioral conformity within the group. The reason is that the individual may feel that deviations from group behavior, whether positive or negative, signal a lack of commitment to the group, which may trigger disapproval by other group members. Thus, the individual will be more vigilant — and responsive — to both examples of compliance and transgression.

To study this conjecture in our behavioral experiment, we designed a treatment where participants could observe not only other subjects' behavior but also a minimal cue about social proximity with these subjects (fandom of a sports team). Social proximity significantly reduced the asymmetry in reactions to observing norm violations and norm compliance. Under social proximity, people not only paid attention to norm violations but were also willing to emulate compliant behavior, thus halting the erosion of norm compliance. Our results about the importance of social proximity for norm compliance are in line with a large literature that shows significant differences in behavior towards socially proximate as opposed to socially distant others ("in-groups" vs. "out-groups"). For instance, individuals are more altruistic, trusting, and trustworthy towards in-groups rather than out-groups (McEvily et al., 2006; Balliet et al., 2014), more likely to comply with requests by an in-group rather than an out-group member (Burger et al., 2004), and more likely to take an in-group's advice into account (Gino et al., 2009b, and see Lane (2016) for a review and meta-analysis of this literature.).

Our novel contribution is showing that social proximity can stabilize norm compliance by reducing the asymmetry in reactions to observing norm violations and norm conformity. Punishment, often seen as an important stabilizer of social norms, may not be necessary when social proximity induces emulation of norm compliance. This highlights the importance of integrating the broader social context in the study of norm compliance. Most existing experimental research studies norms in abstract, anonymous, and context-neutral decision settings. While using contextually neutral decision environments is one of the hallmarks of experimental control, we show that this comes at the cost of missing important insights about the drivers of norm compliance: our results show that one would draw

substantially different conclusions about the effects of observation on norm compliance depending on whether interactions occur among strangers or socially proximate individuals.

Our results have also key implications for the design of behavioral change interventions (for a discussion, see Bicchieri and Dimant, 2019). We show that providing information about the behavior of anonymous strangers — an intervention that is commonly used by designers of behavioral change — is likely to backfire when the information reveals that non-compliance is widespread. This is consistent with previous studies finding limited effectiveness for this type of interventions (e.g., Cialdini et al., 2006; Dimant et al., 2020). Former research has discussed mechanisms that could counterbalance this negative effect of peer information, e.g., by providing individuals with a reminder of the normative value of compliance (Schultz et al., 2007). Our findings suggest an alternative solution: providing individuals targeted information about the behavior of *similar* others may be sufficient to halt the erosion of norm compliance.

References

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of personality and social psychology*, 63(4):596.

Azrieli, Y., Chambers, C. P., and Healy, P. J. (2018). Incentives in experiments: A theoretical analysis. *Journal of Political Economy*, 126(4):1472–1503.

Balliet, D., Wu, J., and De Dreu, C. K. (2014). Ingroup favoritism in cooperation: A meta-analysis. *Psychological Bulletin*, 140(6):1556.

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

Barr, A., Lane, T., and Nosenzo, D. (2018). On the social inappropriateness of discrimination. *Journal of Public Economics*, 164:153–164.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bicchieri, C. and Chavez, A. (2010). Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making*, 23(2):161–178.

Bicchieri, C. and Dimant, E. (2019). Nudging with care: The risks and benefits of social information. *Public Choice*.

Bicchieri, C., Dimant, E., and Sonderegger, S. (2020a). It's not a lie if you believe the norm does not apply: Conditional norm-following with strategic beliefs. Working Paper Available at SSRN: https://dx.doi.org/10.2139/ssrn.3326146.

Bicchieri, C., Dimant, E., and Xiao, E. (2020b). Deviant or wrong? The effects of norm information on the efficacy of punishment. Working Paper Available at SSRN: https://dx.doi.org/10.2139/ssrn.3294371.

Bolton, G., Dimant, E., and Schmidt, U. (2019). The fragility of nudging: Combining (im)plausible deniability with social and economic incentives to promote behavioral change. Working Paper Available at SSRN: https://dx.doi.org/10.2139/ssrn.3294375.

Bond, R. (2005). Group size and conformity. *Group Processes & Intergroup Relations*, 8(4):331–354.

Burger, J. M., Messian, N., Patel, S., del Prado, A., and Anderson, C. (2004). What a coincidence! The effects of incidental similarity on compliance. *Personality and Social Psychology Bulletin*, 30(1):35–43.

Bursztyn, L., Egorov, G., and Fiorin, S. (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review*.

Chang, D., Chen, R., and Krupka, E. (2019). Rhetoric matters: A social norms explanation for the anomaly of framing. *Games and Economic Behavior*, 116:158–178.

Charness, G., Gneezy, U., and Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131:141–150.

Charness, G., Naef, M., and Sontuoso, A. (2019). Opportunistic conformism. *Journal of Economic Theory*, 180:100–134.

Chen, Y. and Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, 99(1):431–57.

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., and Winter, P. L. (2006). Managing social norms for persuasive impact. *Social influence*, 1(1):3–15.

Coleman, J. S. (1994). *Foundations of social theory*. Harvard University Press.

d'Adda, G., Drouvelis, M., and Nosenzo, D. (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, 62:1–7.

d'Adda, G., Dufwenberg, M., Passarelli, F., and Tabellini, G. (2020). Social norms with private values: theory and experiments. *Games and Economic Behavior*, 124:288–304.

Depetris-Chauvin, E., Durante, R., and Campante, F. R. (2020). Building nations through shared experiences: Evidence from african football. *American Economic Review*.

Dimant, E. (2019). Contagion of pro-and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, 73:66–88.

Dimant, E. (2020). Hate trumps love: The impact of political polarization on social preferences. Working Paper Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3680871.

Dimant, E., Gerben, A. v. K., and Shalvi, S. (2020). Requiem for a nudge: Framing effects in nudging honest. *Journal of Economic Behavior & Organization*, 172:247–266.

Dimant, E. and Gesche, T. (2020). Nudging enforcers: How norm perceptions and motives for lying shape sanctions. Working Paper Available at SSRN: https://dx.doi.org/10.2139/ssrn.3664995.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Eckel, C. C. and Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and Economic Behavior*, 16(2):181–191.

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

Fehr, E. and Schurtenberger, I. (2018). Normative foundations of human cooperation. *Nature Human Behaviour*, 2(7):458.

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

Fischbacher, U. and Gächter, S. (2010). Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *American Economic Review*, 100(1):541–56.

Fisman, R. and Miguel, E. (2007). Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets. *Journal of Political Economy*, 115(6):1020–1048.

Gächter, S., Gerhards, L., and Nosenzo, D. (2017). The importance of peers for compliance with norms of fair sharing. *European Economic Review*, 97:72–86.

Gächter, S., Molleman, L., and Nosenzo, D. (2020). The behavioral logic of rule following and social norm compliance. Unpublished Manuscript.

Gächter, S., Starmer, C., and Tufano, F. (2015). Measuring the closeness of relationships: a comprehensive evaluation of the'inclusion of the other in the self'scale. *PloS one*, 10(6):e0129478.

Gangadharan, L., Jain, T., Maitra, P., and Vecci, J. (2016). Social identity and governance: The behavioral response to female leaders. *European Economic Review*, 90:302–325.

Gino, F., Ayal, S., and Ariely, D. (2009a). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science*, 20(3):393–398.

Gino, F., Shang, J., and Croson, R. (2009b). The impact of information from similar or different advisors on judgment. *Organizational Behavior and Human Decision Processes*, 108(2):287–302.

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

Hogg, M. A. (1992). *The social psychology of group cohesiveness: From attraction to social identity.* Harvester Wheatsheaf.

Keizer, K., Lindenberg, S., and Steg, L. (2008). The spreading of disorder. *Science*, 322(5908):1681–1685.

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

Kimbrough, E. O. and Vostroknutov, A. (2018). A portable method of eliciting respect for social norms. *Economics Letters*, 168:147–150.

Kocher, M. G., Schudy, S., and Spantig, L. (2017). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science*, 64(9):3995–4008.

Korenok, O., Millner, E. L., and Razzolini, L. (2014). Taking, giving, and impure altruism in dictator games. *Experimental Economics*, 17(3):488–500.

Krupka, E. L. and Croson, R. T. (2016). The differential impact of social norms cues on charitable contributions. *Journal of Economic Behavior & Organization*, 128:149–158.

Krupka, E. L., Leider, S., and Jiang, M. (2017). A meeting of the minds: informal agreements and social norms. *Management Science*, 63(6):1708–1729.

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

Lane, T. (2016). Discrimination in the laboratory: A meta-analysis of economics experiments. *European Economic Review*, 90:375–402.

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

McEvily, B., Weber, R. A., Bicchieri, C., and Ho, V. T. (2006). Can groups be trusted? an experimental study of trust in collective entities. *Handbook of Trust Research*, pages 52–67.

Mussweiler, T. (2001). 'seek and ye shall find': Antecedents of assimilation and contrast in social comparison. *European journal of social psychology*, 31(5):499–509.

Mussweiler, T. (2003). Comparison processes in social judgment: mechanisms and consequences. *Psychological review*, 110(3):472.

Perkins, H. W. (2002). Social norms and the prevention of alcohol misuse in collegiate contexts. *Journal of Studies on Alcohol, supplement*, (14):164–172.

Reuben, E. and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1):122–137.

Schram, A. and Charness, G. (2015). Inducing social norms in laboratory allocation choices. *Management Science*, 61(7):1531–1546.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429–434.

Strathman, A., Gleicher, F., Boninger, D. S., and Edwards, C. S. (1994). The consideration of future consequences: weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology*, 66(4):742.

Sutter, M., Haigner, S., and Kocher, M. G. (2010). Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *The Review of Economic Studies*, 77(4):1540–1566.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, 33(1):1–39.

Tajfel, H. and Turner, J. C. (1979). An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations*, 33(47):56–65.

Tangney, J. P., Baumeister, R. F., and Boone, A. L. (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality*, 72(2):271–324.

Thöni, C. and Gächter, S. (2015). Peer effects and social preferences in voluntary cooperation: A theoretical and experimental analysis. *Journal of Economic Psychology*, 48:72–88.

Online Appendix to "Social Proximity and the Erosion of Norm Compliance"

Cristina Bicchieri, Eugen Dimant, Simon Gächter, Daniele Nosenzo

Contents

I.a. Robustness Checks (Experiment 1)

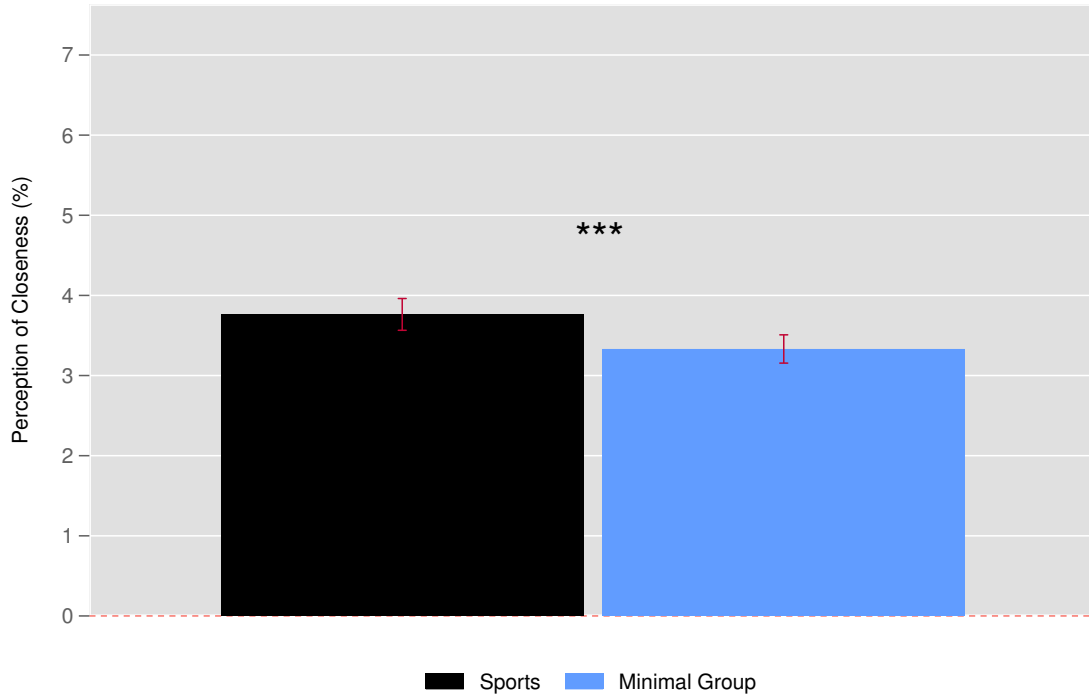**I.a.1. Perceived Closeness Using Sports Identity vs. Minimal Group Identity**

Figure A.1: Comparison between sports identity and minimal group identity using the IoS scale. The difference is significant at the 1% level using a two-sided MWU test. For more information see Footnote 5.

I.a.2. Additional Analyses of Norm Compliance Erosion

Here we present an additional analysis to examine the erosion of norm compliance for those subgroups that have not started with a majority of taking behavior. Our findings strongly support our main results: compared to NoObservation, even when taking behavior is in the minority from the beginning, the erosion of norm compliance accelerates in the Observation condition (p=0.0201), while being substantially muted in the presence of social proximity in ObservationSP (p=0.8731 compared to NoObservation; p<0.01 compared to Observation).
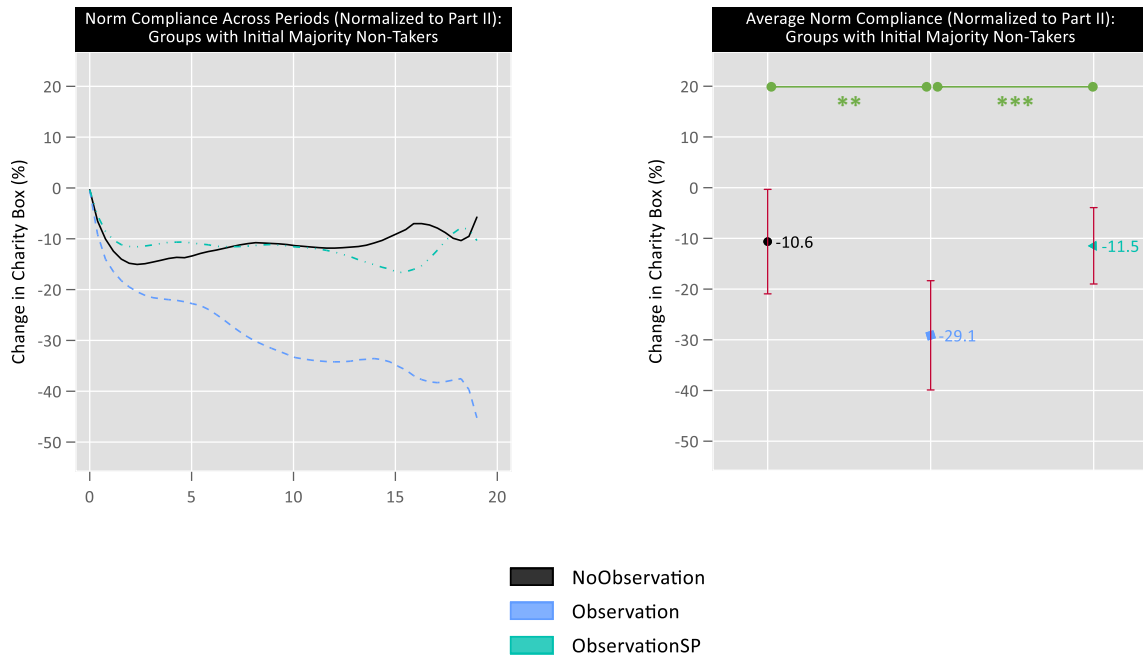
Figure A.2: Left panel: Subject's behavior across treatments and periods in Part III for groups in which the initial average behavior was not taking, normalized to average behavior in Part II (period 0 in the graph). For purposes of readability, lines represent kernel-weighted local polynomial smoothing of degree 5. Right panel: Average subject's donations across treatments, averaged over the 19 periods of Part III. Whiskers indicate 95% confidence intervals. Stars indicate significant differences using two-sided Mann-Whitney U-tests and group averages as unit of observation, with p-values adjusted for multiple hypothesis testing (Benjamini and Hochberg, 1995). *p<0.1, **p<0.05, and ***p<0.01.

I.a.3 Group Proximity

Here we present graphical illustrations of the evolution of norm compliance in ObservationSP. Figure A.3 presents this for homogeneous groups, where all group members answered the proximity question in the same way (either all correctly or all incorrectly), as well as for heterogeneous groups (mixed answers within the same group). Figure A.4 breaks down the data for the homogeneous groups in one additional way: groups where all answers to the proximity questions were correct compared to groups where all answers were incorrect. Though we did not have enough data to draw definitive conclusions, we observe a tendency for groups where the similarity is strong *and* significant (all are Phillies fans) to respond much more strongly to norm abiding behavior from their peers.
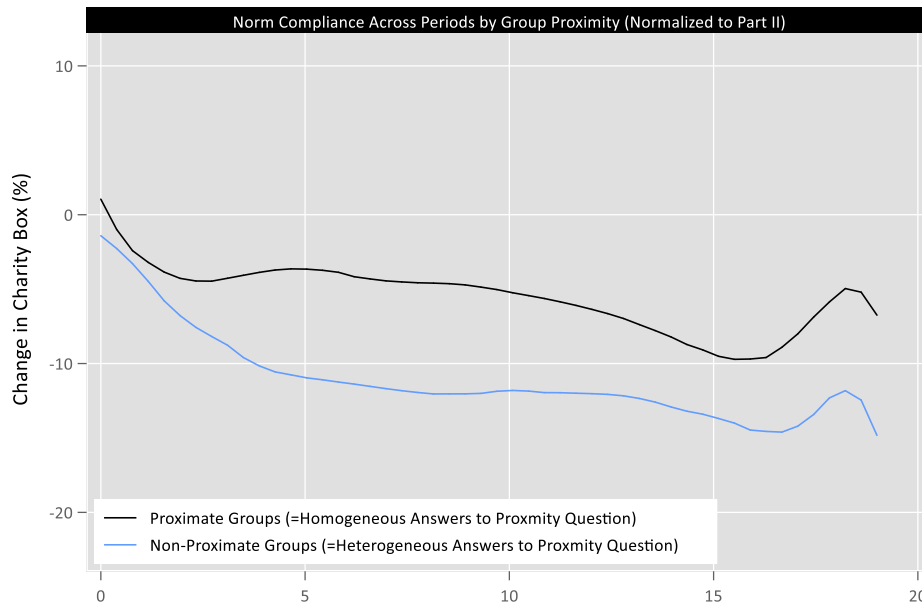
Figure A.3: Evolution of norm compliance (normalized to Part II behavior) in ObservationSP for groups answering the proximity question in the same way (proximate) or in different ways (non-proximate). Polynomial smoothing applied.
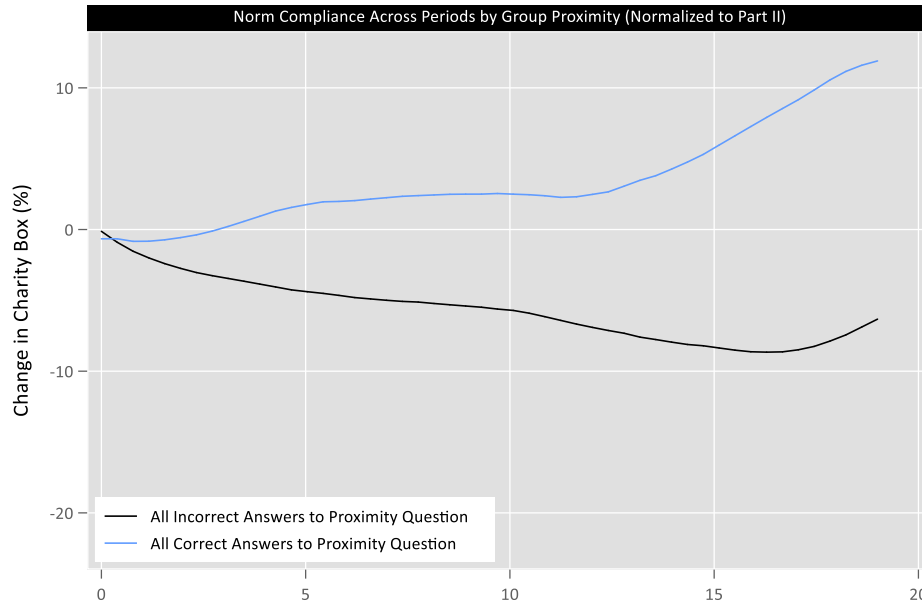

Figure A.4: Evolution of norm compliance (normalized to Part II behavior) in ObservationSP for groups answering the proximity question correctly or incorrectly. Polynomial smoothing applied.

4

I.a.4 Additional Heterogeneity Analysis

The figure below presents a more detailed breakdown of our Figure 3 in that it subdivides the behavioral response of initial norm violators and norm followers along two dimension: took a small amount from/gave a small amount to the charity (left panel) versus took a large amount from/gave a large amount to the charity (right panel). Our main results for the initial norm followers remain robust showing that — irrespective of the breakdown — Observation yields the largest erosion of norm compliance, whereas both NoObservation and ObservationSP indicate lower levels of erosion that are statistically indistinguishable from each other. On top of that, however, these results also suggest that the initial takers do in fact show heterogeneous treatment effects depending on whether they initially took a lot or a little. This suggests that even initial takers can be swayed by their peers, although mainly accelerating erosion of norm compliance if their peers behaved worse than they did.
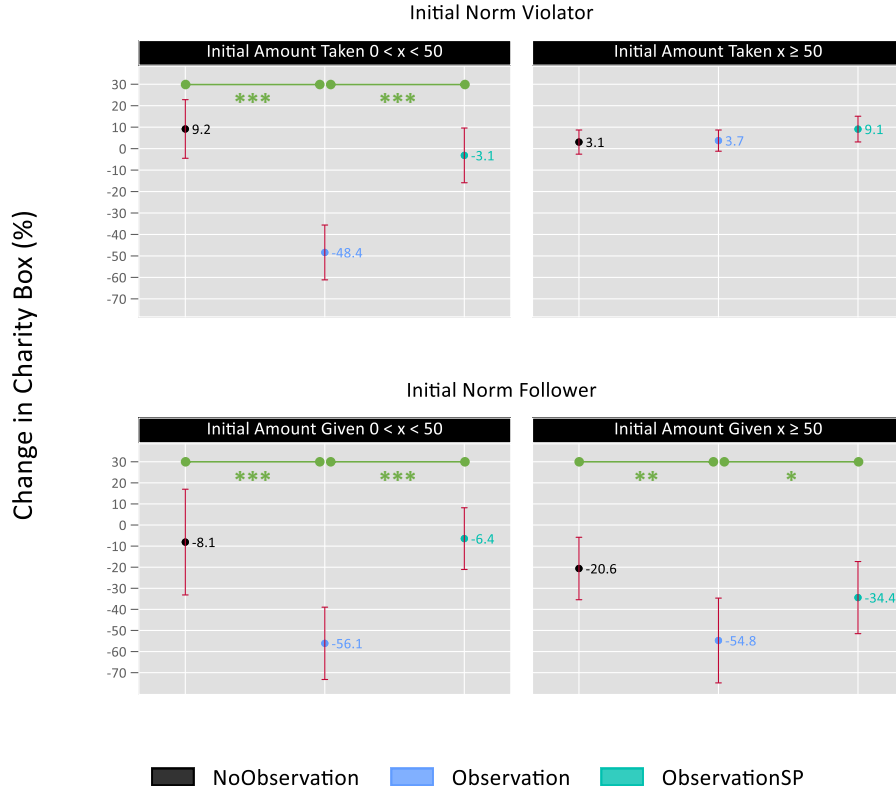

Figure A.5: Average subject's donations across treatments, averaged over the 19 periods of Part III conditional on their initial behavior in Part II and how extreme that behavior was (cut-off at +50 and -50, respectively). Whiskers indicate 95% confidence intervals. Stars indicate significant differences at the conventional levels of *p<0.1, **p<0.05, and ***p<0.01.

We can further substantiate our results from Figure 3 in the main text by presenting a more detailed breakdown of changes in norm compliance across treatments and individual types. In line with the previously discussed results, the initial norm followers display the largest changes in behavior, followed by initial abstainers and initial norm violators. We present the results in Figure A.6. In addition, we can examine a similar breakdown for both initial norm violators and initial norm followers based on how extreme their Part II behavior was (see Figure A.5). Note: the results do not change whether the cut-off values of -50 and +50 are included in the lower or upper half.
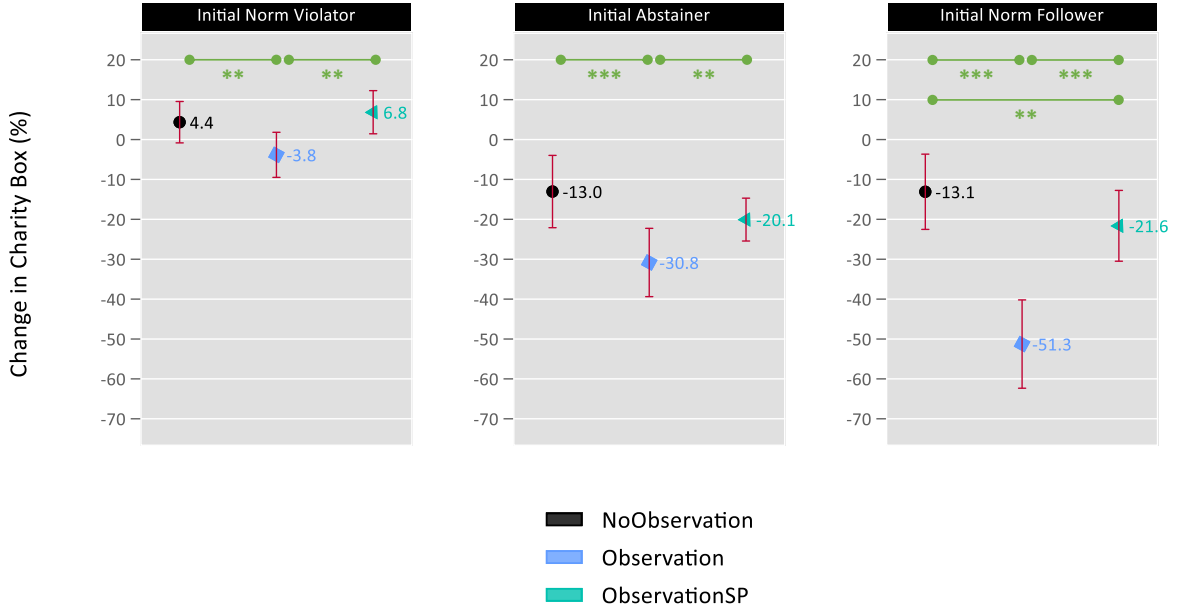

Figure A.6: Average subject's donations across treatments, averaged over the 19 periods of Part III conditional on their initial behavior in Part II. Whiskers indicate 95% confidence intervals. Stars indicate significant differences at the conventional levels of *p<0.1, **p<0.05, and ***p<0.01.

I.b. Instructions (Experiment 1)

Below, we present the instructions for all our treatment variations. Recall that the instructions for the first decision were the same across all treatments and only the instructions for the subsequent group stage differed. We highlight the text in green *that was used instead of the preceding numbers and sentences between the NoObservation and the Observation and ObservationSP treatments. For the latter, the instructions were the same. Information about the social proximity was not pre-announced but simply displayed on the computer screen (see screenshots in Appendix III).*

Instructions for First Decision

General

Thank you for coming! You have earned $10 for showing up on time. The following instructions explain how you can potentially earn more money by making a number of decisions. To maximize your chances to earn more money, please read these instructions carefully! If you have a question at any time, please raise your hand, and an experimenter will assist you. For the purpose of the experiment, it is important that you do not talk or communicate in other ways with the other participants. Please turn off or silence your cell phone and all other electronic devices. You are asked to abide by these rules. If you do not abide, we will have to exclude you from this and future experiments (with the current experimenter) and you will not receive any compensation for the current experiment. Your decisions will remain anonymous to any other participants throughout the experiment. No participant will know who has made what decisions. Please do not talk to each other during the experiment. During the experiment, all amounts will be presented in ECU (Experimental Currency Units). At the end of the experiment, the ECU you have earned will be converted to Dollars as follows:

$$10 \text{ ECU} = 1 \text{ Dollar}$$

Roles in the Experiment

- There are two roles in the experiment:

  1. Decision Maker
  2. Receiver

- Each participant in the experiment plays in the role of the decision maker.

7

- The role of the receiver is represented by one of the following three charities:

  1. Doctors Without Borders
  2. World Wildlife Fund (WWF)
  3. UNICEF

Order of Events

- You will make multiple decisions in this experiment. Explanations and information related to these decisions will be given at the relevant points throughout the experiment.

- Both you as well as the receiver in the form of a charitable organization will be provisionally assigned a monetary amount of 100 ECU each.

- The decision maker will have to decide whether to. . .

  - . . . take a part or all of the money from the receiver.
  - . . . leave the division of the sum of money as it is.
  - . . . give a part or all of his/her money to the receiver.

- At the end of the experiment, one decision maker will be chosen at random, exactly one decision of this decision maker will be implemented, and count towards his/her own payoff as well as the receiver. The randomly selected participant at the end of the experiment will determine the receiving charity.

  - The decision of every other decision maker will not count towards his or her own and the receiver's payoff. Instead, each of these decision makers will receive exactly 100 ECU (their show-up fee) regardless of their allocation decision towards the receiver.

- The ECU amount the receiving charity is left with at the end will be doubled (meaning that the multiplier equals 2) and actually transferred to the respective charity. We will upload a receipt of the donation to our website (http://www.sas.upenn.edu/ppe/) where you can verify the validity of the donation.

8

Payoff Structure

- The total payoff of the participants:

  - **If you are the randomly chosen decision maker**

    * 100 ECU (show-up fee) +/- the amount of money that was given to/taken from the receiver

  - **If you are not the randomly chosen decision maker**

    * 100 ECU (show-up fee)

- The total payoff of the receiving charity:

  - (100 ECU +/- amount of money that was given to/taken from the charity by the randomly chosen decision maker) x2

9

Instructions for Group Stage (Distributed After Initial Decision)

We highlight the treatment differences in *green*

General

- Every decision maker is now randomly paired up with three other decision makers in this room. That is, each group consists of exactly 4 (2) decision makers.

- This grouping will not change over the course of the next 19 rounds, meaning you will remain paired up with the same 3 (1) decision makers from now until the end of the experiment. The experiment ends after round 20.

- Over the course of the next 19 rounds, every decision maker continues to make the same decisions as before. That is, the decision maker will have to decide whether to. . .

  - . . . take a part or all of the money from the receiver.
  - . . . leave the division of the sum of money as it is.
  - . . . give a part or all of his/her money to the receiver.

- The decisions that are made over the course of the next 19 periods will remain anonymous. This means that individual behavior will not be displayed to any participant within your group or anyone else. (At the end of each round, every decision maker's decision will be revealed to every other decision maker inside the group. The decisions being made over the course of the next 19 rounds will be displayed within each group, but not to anyone else.)
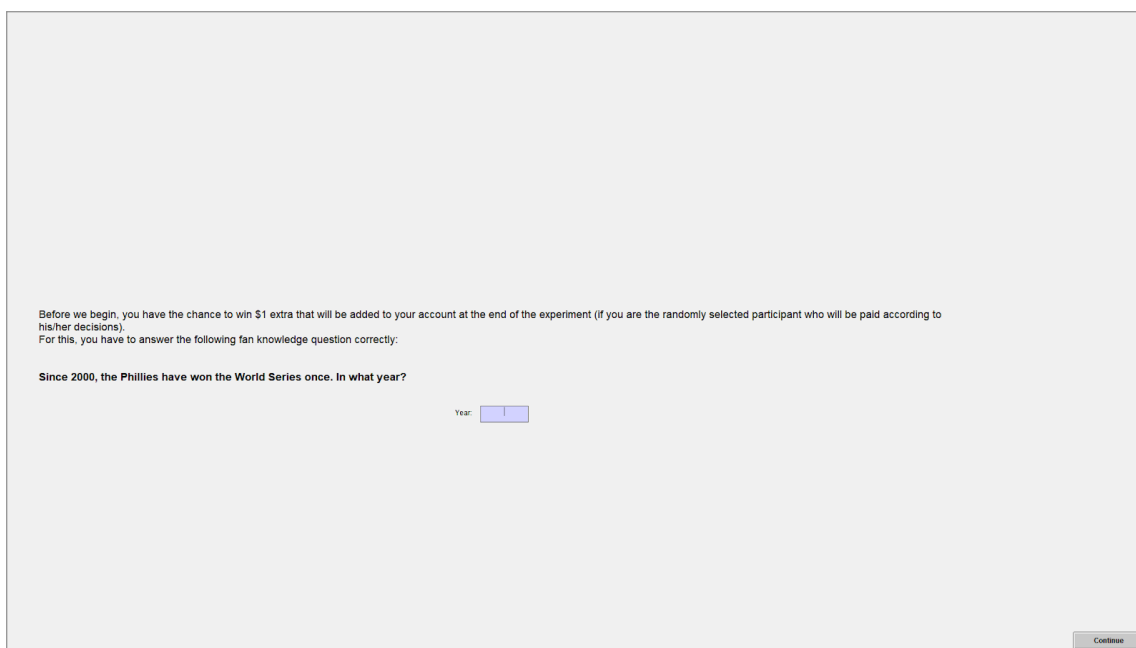
Payoff

- The payoff mechanism for this part of the experiment is exactly the same as for the first part of the experiment. That is (as a reminder):

  - At the end of the experiment, one decision maker will be chosen at random, exactly one decision of this decision maker will be implemented, and count towards his/her own payoff as well as the receiving donation.
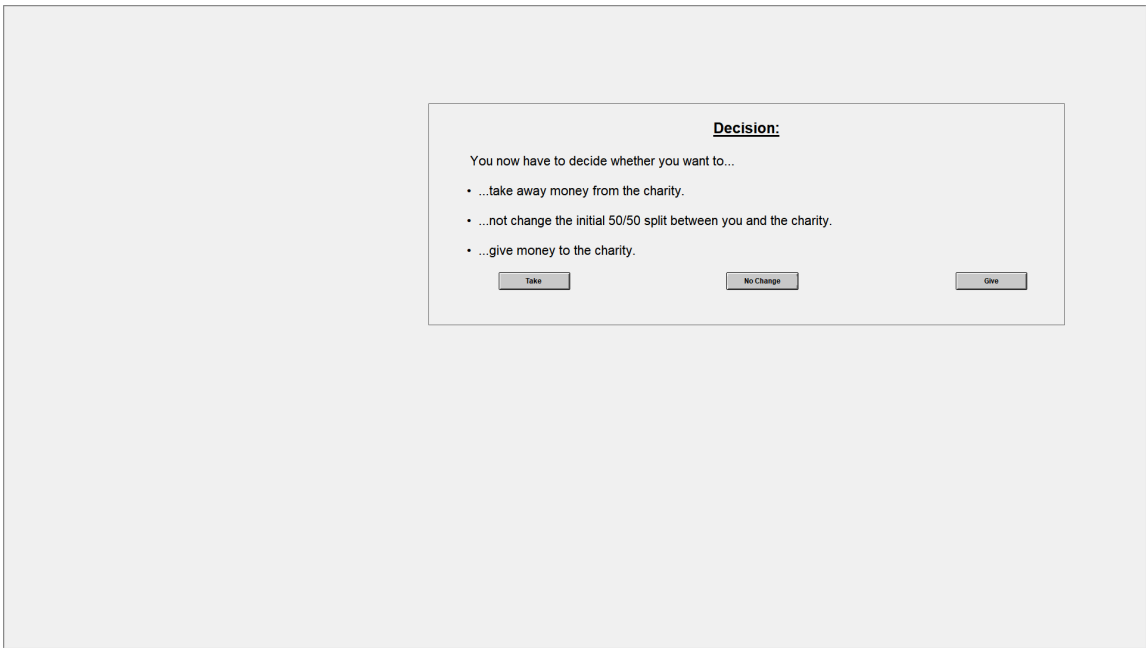
10

- The decisions of every other decision maker will not count towards his or her own and the receiver's payoff. Instead, each of these decision makers will receive exactly 100 ECU regardless of their allocation decision towards the receiver.

- The final ECU amount left with the receiving charity will be doubled and actually transferred to the respective charity. We will upload a receipt of the donation to our PPE website (http://www.sas.upenn.edu/ppe/) where you can verify the validity of the donation.

- The decisions that are made over the course of the next 19 rounds will remain anonymous. This means that individual behavior will not be displayed to any participant within your group or anyone else. (At the end of each round, every decision maker's decision will be revealed to every other decision maker inside the group. The decisions being made over the course of the next 19 rounds will be displayed within each group, but not to anyone else.)
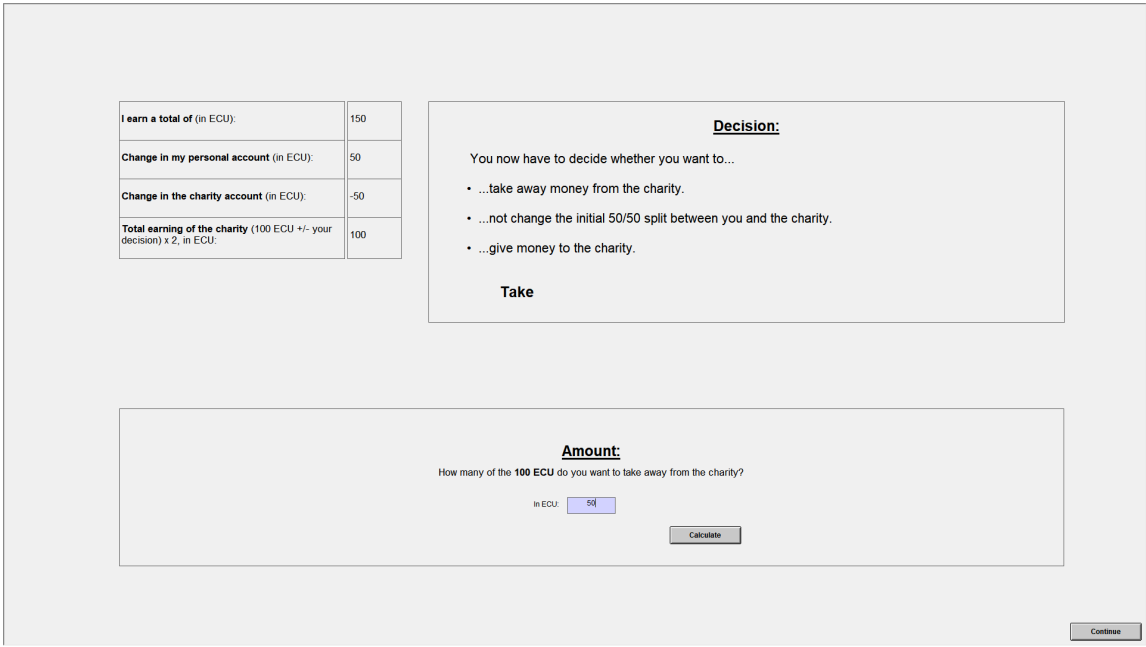
11

I.c. Experimental Screenshots (Experiment 1)

Here, we present the screenshots for ObservationSP with a group size of 4. Screenshots for other treatments are available upon request. Merely Figure A5 changed across the different conditions: in NoObservation, participants did not see any information in other group members' history table. In the Observation condition, participants observed the behavior of all participants in their group for all periods. For ObservationSP (as displayed in Figure A5), participants observed the behavior of all participants in their group for all periods plus the information as to whether or not the respective participant answered the knowledge question correctly.

Before we begin, you have the chance to win $1 extra that will be added to your account at the end of the experiment (if you are the randomly selected participant who will be paid according to his/her decisions).
For this, you have to answer the following fan knowledge question correctly:

Since 2000, the Phillies have won the World Series once. In what year?

Year: [ ]

Continue

Screen 1: Knowledge question (correct answer: 2008).

Screen 2a: Decision towards charity (choosing the type of behavior).

Screen 2b: Decision towards charity (choosing the amount).

13

Screen 3: Observation stage (after 3 periods).

Screen 4: Observation stage (after 3 periods).

14

You were randomly chosen and your decision towards the charity will be implemented.

Your payoff in ECU is:	160
Your payoff in Dollar is:	16
Out of which your additional payoff for correctly answering the knowledge question Dollar is:	1
The randomly chosen round is:	1
Total payoff of the receiving charity in Dollar:	15

Continue

Screen 5: Final payoff screen (if participant was randomly chosen).

You were not randomly chosen and your decision towards the charity will not be implemented.

Hence, your payoff in this experiment is **10 Dollars**.

Continue
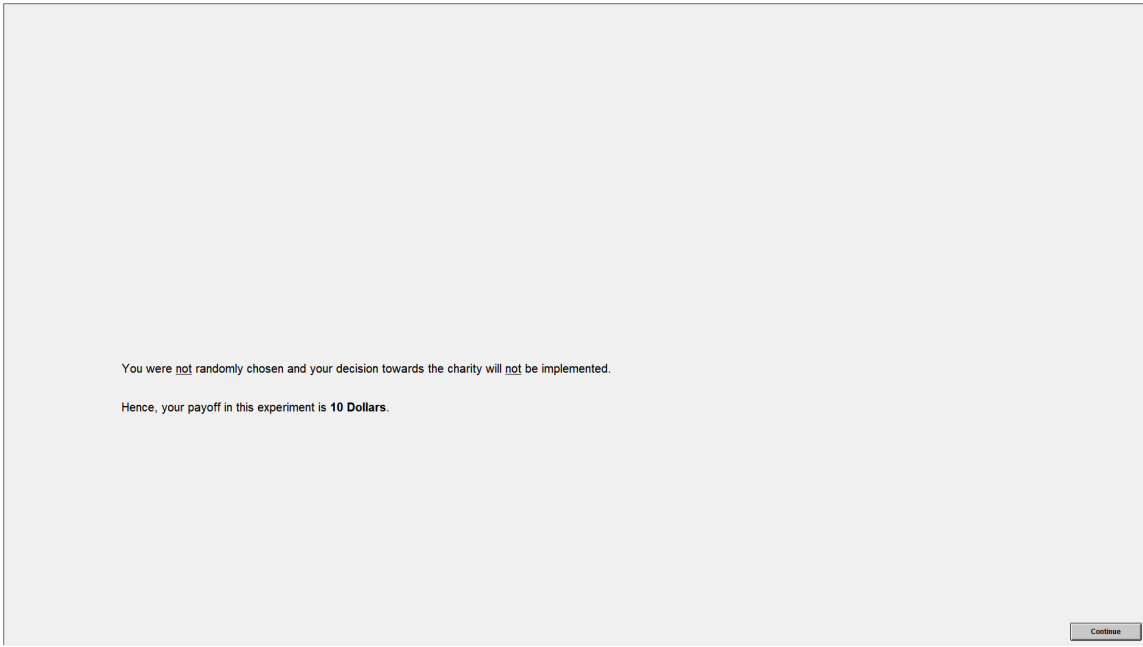
Screen 6: Final payoff screen (if participant was not randomly chosen).

15

II.a. Design (Experiment 2)

We designed one separate experiment to reconstruct whether a norm of giving to charity exists in our ToG game. Conceptually, we build on Bicchieri (2006, 2016). In her theory of social norms, three types of beliefs matter for norm compliance: (1) personal beliefs of appropriateness, (2) normative expectations, and (3) empirical expectations. Personal beliefs are first-order beliefs that capture the individual's own opinion about what constitutes appropriate or inappropriate behavior in a given situation. Normative expectations are instead second-order beliefs that describe an individual's beliefs about what others consider appropriate or inappropriate. Finally, empirical expectations capture what the individual believes to constitute common behavior in the situation (i.e., what most other people do). We conducted a norm-elicitation experiment to measure these three types of expectations about behavior in the ToG game. Consistent with our main behavioral experiment, this norm-elicitation experiment was conducted with students from the University of Pennsylvania who had *not* previously participated in any of our other experiments.[17]

To measure beliefs, we used a variant of the procedure introduced by Bicchieri and Chavez (2010). We recruited $N = 110$ Penn students and described the ToG donation game to them. Each subject was asked three types of questions: first, what they personally thought that one should do in the ToG game, which they answered by selecting one of the 21 actions available to a decision-maker in the game. This question measures subjects' *personal beliefs* about what is appropriate in the game. Next, participants were presented, in random order, a question about either their normative or empirical expectations. In the former case, subjects were asked to guess the most common response to the personal belief question by other participants in the same experiment. For empirical expectations, subjects were asked to guess the most frequent choice that subjects actually made in the behavioral experiment conducted with a separate group of subjects. Normative and empirical expectations were incentivized: for both, a correct response yielded a \$1 bonus in addition to a \$2 show-up fee. The average duration of the experiment was 10 minutes.

---

[17]As customary in the literature, we elicited these expectations from uninvolved third parties to mitigate the challenges that arise from eliciting beliefs as part of the same experiment, either when done at the beginning of the experiment (it can introduce noise in the subsequent decision-making, e.g., through priming, demand effects, etc.) or at the end of the experiment (it can give rise to post-hoc justifications). See d'Adda et al. (2016) for further discussion. Since we collect the relevant norm information from exactly the same pool of students who also participated in the main experiment, it is reasonable to expect that the norm-perceptions are sufficiently similar to each other.

II.b. Results (Experiment 2)

Figure A.7 shows the results of the norm-elicitation experiment. Although in the experiment we asked subjects to report their beliefs by selecting one of the 21 possible actions of the ToG game, for ease of presentation, we collapse responses as follows: actions implying taking from the charity ("Take"), actions leaving the charity's endowment untouched ("Abstain"), and actions implying giving to the charity ("Give").
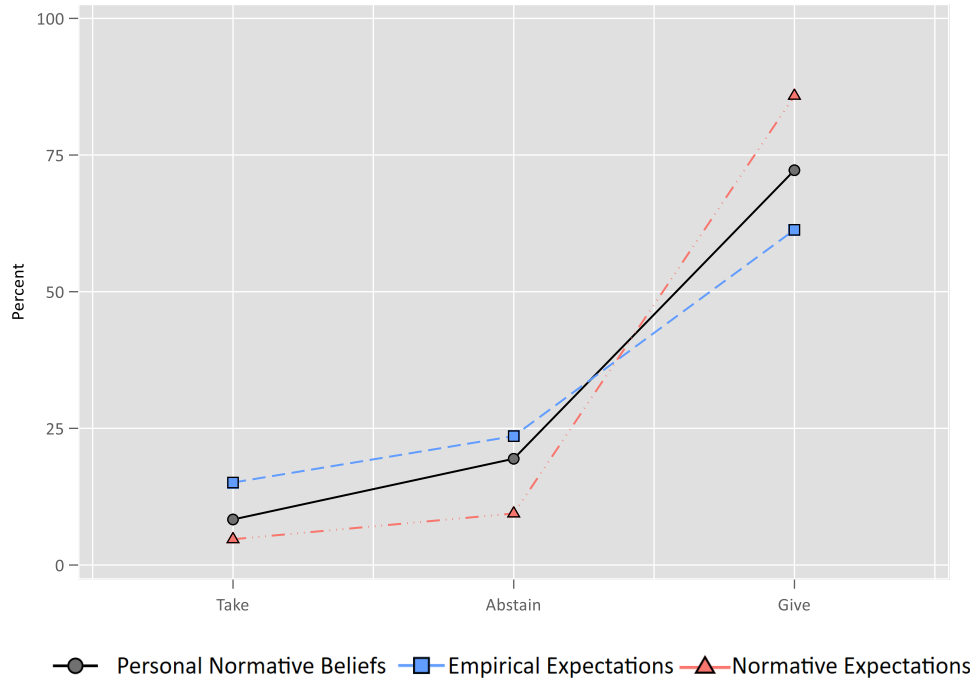

Figure A.7: Personal normative beliefs, normative expectations & empirical expectations in the ToG game.

Subjects' personal beliefs are shown by the black solid line in Figure A.7, representing the percentage of subjects ($N = 108$) who personally believe that one ought to take, give, or abstain from changing endowments in the ToG game. The results in this figure undoubtedly indicate that in our context, most subjects think that one ought to give. The dashed red line in Figure A.7 shows subjects' normative expectations (second-order beliefs about what others believe ought to be done in the game, $N = 106$). The line indicates that second-order beliefs closely track the corresponding (first-order) personal beliefs of prescribed behavior: 85.9% of subjects believe that a majority of others indicated that giving is what one ought to do in the game, and only 4.7% believe that the majority of others thought that one ought to take. The corresponding figures are 72.2% and 8.3%,

respectively, for personal beliefs. Together, the personal beliefs and normative expectations indicate that there are mutually consistent (and correct) expectations among a majority of subjects that giving constitutes prescribed behavior in the experiment. This result is in line with our conjecture that normative expectations that prescribe giving and proscribe taking exist in the ToG game. The fact that they are also consistent with personal beliefs tells us that there is no pluralistic ignorance, i.e., the norm is also personally endorsed by a majority of people.

Finally, the dashed blue line in Figure A.7 represent subjects' empirical expectations ($N = 106$). Note that these are expectations that subjects form in the absence of any feedback information about behavior of others in ToG game, i.e., their initial expectations without any experience with the game. We find that 61.3% of subjects expect a majority of participants in the game to give to charity, and another 23.6% to leave initial endowments unchanged. Only 15.1% of subjects expect others to take in the game. Thus, empirical expectations align with normative expectations: both indicate that giving is what one should do and what most are expected to do. However, it should be noted that empirical expectations are lower than the normative ones.[18] This indicates that there is more skepticism about how many people follow a shared and commonly endorsed rule. This is not surprising, as social norms exist precisely to curb selfish motivations that are always present in situations where compliance has a cost.

---

[18]A Stuart-Maxwell $\tilde{\chi}^2$ test shows that the distributions are significantly different at the 1% level. In addition, a Wilcoxon matched-pair signrank test that compares the percentage of people who said that they expect a majority of participants to give to charity to those who said that they expect a majority to approve of giving to charity indicates significance at the 1% level, too. For a detailed breakdown of the distribution of beliefs across conditions, see Figure A.8.
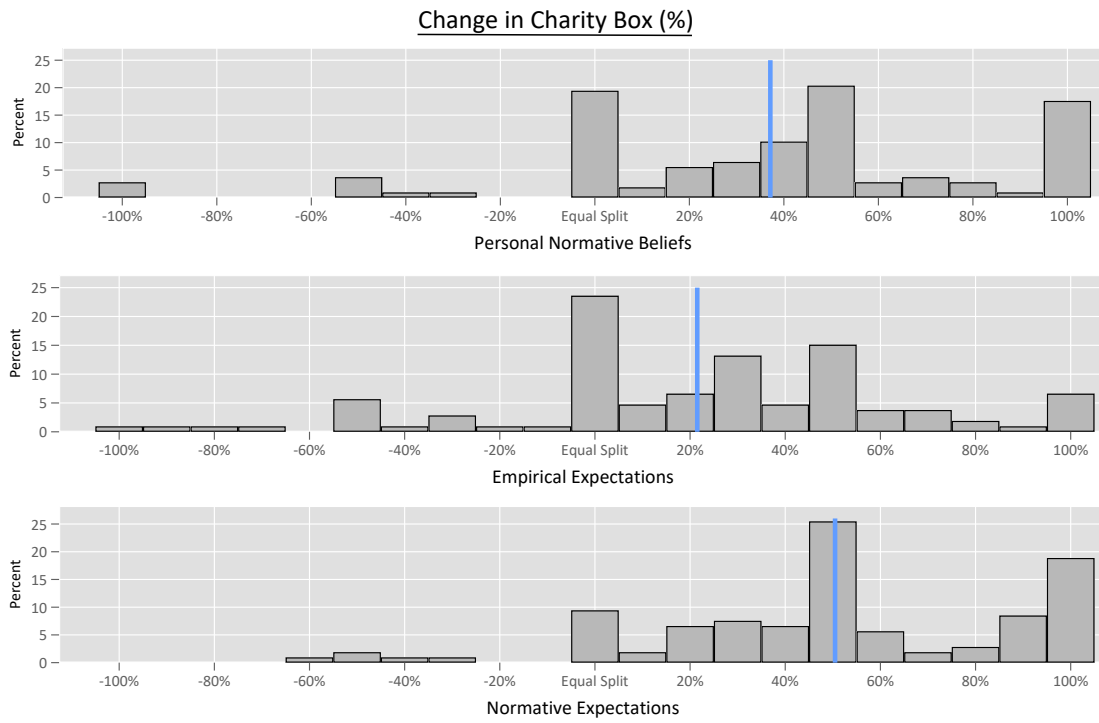
Figure A.8: Distribution of elicited beliefs across conditions. Vertical blue lines indicate averages.

II.c. Experimental Screenshots (Experiment 2)

Consent Form & General Instructions

University of Pennsylvania
Center for Social Norms and Behavioral Dynamics
Department of Philosophy, Politics, and Economics
Claudia Cohen Hall, Room 311
Philadelphia, PA 19104

Phone: (215)-898-3023
Fax: (215)-573-2231

Informed Consent Form

You are invited to take part in a study named *Guessing Task*. The purpose of this research study is to understand human decision-making. You will answer a series of questions regarding beliefs and behaviors. We will also ask you to provide demographic information. We will not ask for your name or any information that will make you identifiable. Overall, this study will take approximately 5-10 minutes.

For your participation in this study, you will receive a fixed payment of $2. Additionally, you may receive a monetary bonus. If a question is eligible for a monetary bonus, it will clearly state so.

You will be paid within 7 days of completing the study through an Amazon gift card delivered to the email address you will provide at the end of this study. Your email address will only be used to pay you correctly and will be deleted permanently from the experimenter's data after payment is complete.

The risks to participating are no greater than those encountered in everyday life. Your participation in this study is completely voluntary, and you may refuse to participate. If you withdraw from the study before completing it, you will not be paid. Compensation will be awarded upon completion of the entire study.

If you have any questions about this study, you may contact the Behavioral Ethics Lab at behavioralethicslab@gmail.com.

For any questions, concerns, suggestions, or complaints that are not being addressed by the researcher, please contact the Institutional Review Board at the University of Pennsylvania, 3624 Market Street, Suite 301 South Philadelphia, PA 19104-6006. Phone: (215) 898-2614.

Please feel free to print or save a copy of this consent form.

By continuing from this page you are indicating that you have read and understood this consent form and wish to continue your participation in this study.

The next button will be enabled after 10 seconds.

Consent

○

General Instructions

Thank you for choosing to participate in this study! It is important that you **read all of the instructions carefully to maximize your earnings**. At the end of the study, you will be asked to fill out a short questionnaire and demographic questions.

It is expected that this study will take approximately 5-10 minutes. There is **no deception** in this study. Everything you see or read is true.

You will be paid $2 for participation and will have the opportunity to earn more based on some of your answers. For example, you will be asked to guess what other participants have previously done or thought should be done. Guessing correctly will earn you an additional $1 for each question that is eligible for payment.

Scenario

In what follows, we describe a scenario based on a previous experiment.
The experiment involved real stakes and participants' decisions were actually implemented.

In the experiment, the participant and a charity started with an endowment of $10 each. The charity was chosen by the participant from the following three: Doctors without Borders, World World Wildlife Fund (WWF), and UNICEF.

The participant had to decide whether to:

- ...**take** a part or all of the money **from the charity**
- ...**make no change** to either of the endowments
- ...**give** a part or all of the their money **to the charity**

For example, making no change means choosing 0 on the following slider. Giving to a charity means choosing a positive amount on the slider. Taking from the charity means choosing a negative amount on the slider.

| Taking | | | | | | | | | | No change | | | | | | | | | Giving |
| -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Amount ($)

Below, we also give 3 examples showing people taking, giving, and making no change to the charity's endowment.

| Taking | | | | | | | | | | No change | | | | | | | | | Giving |
| -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Example: Taking $5 from the charity

Example: Giving $2 to the charity

Example: Making no change to the charity

21

Three Belief Elicitations (Within-Design)

1. Personal Normative Beliefs (PNB) - Decision

Based on the scenario described above, please mark on the slider **what you think one should do**

Taking										No change										Giving
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10

Amount ($)

2. Normative Expectations (NE) - Decision

Now, you will make a payoff relevant guess. You will earn an additional $1 if your guess is correct or if it is one unit above or below the correct answer.

This experiment has already been played. We have data about what participants in this experiment did.

Please guess what you believe the **most frequent choice the participants made** in the experiment about taking, giving, or making no change to the charity.

Taking										No change										Giving
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10

Amount ($)

How confident are you about your choice?

No confidence	Slight confidence	Moderate confidence	High confidence
O	O	O	O

3. Empirical Expectations (EE) – Decision

Now, you will make a payoff relevant guess. You will earn an additional $1 if your guess is correct or if it is one unit above or below the correct answer.

We have asked all participants in this survey what **they believe should be done**. Please guess what you believe is the **most frequent answer** that the participants in this survey give about what they believe should be done.

Taking										No change										Giving
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10

Amount ($)

How confident are you about your choice?

No confidence	Slight confidence	Moderate confidence	High confidence
O	O	O	O

Post-Experimental Questionnaire

What is your gender?

○ Female

○ Male

What is the highest level of school you have completed or the highest degree you have received?

○ Less than high school degree

○ High school graduate (high school diploma or equivalent including GED)

○ Some college but no degree

○ Associate degree in college (2-year)

○ Bachelor's degree in college (4-year)

○ Master's degree

○ Doctoral degree

○ Professional degree (JD, MD)

How old are you in years?

How do you see yourself:

Please indicate on the scale if are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?

0 means: `not at all willing to take risks'
10 means: `very willing to take risks'

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Risk Willingness

○

What is your general opinion of charitable organizations?

'1' = I do not support the work of charitable organizations.
'10' = I fully support the work of charitable organizations.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

General Opinion of Charities

○────────────────────────────────────

What is your general opinion of the United Nations International Children's Emergency Fund (UNICEF)?

'1' = I do not support the work of UNICEF.
'10' = I fully support the work of UNICEF.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

General Opinion of the UNICEF

○────────────────────────────────────

What is your general opinion of the Doctors Without Borders?

'1' = I do not support the work of Doctors Without Borders.
'10' = I fully support the work of Doctors Without Borders.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

General Opinion of the Doctors Without Borders

○────────────────────────────────────

What is your general opinion of the World Wildlife Fund (WWF)?

'1' = I do not support the work of the WWF.
'10' = I fully support the work of the WWF.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

General Opinion of the WWF

○────────────────────────────────────

III.a. Design (Experiment 3)

To test the robustness of our results from Experiment 2, we conducted two additional norm-elicitation experiments (total N from both elicitations = 464).

First, we run a version of Experiment 2 that follows the elicitation by Bicchieri and Chavez (2010), but run using a between-subject design with members of the general population (recruited on Amazon Mechanical Turk), rather than students. In this norm elicitation, three distinct sets of subjects (total N = 312) were asked to report their personal normative beliefs, normative expectations, and empirical expectations. The three sets of elicitations were carried out as follows: (1) We asked one group of subjects ($n = 107$) to report what they thought one "should do" (i.e., their personal normative beliefs) in the ToG game. To make the task more manageable, we divided the actions space in three groups that constitute natural partitions of the action space in the game: take from the charity; give to the charity; or abstain from changing the initial endowments. (2) We then asked another group of subjects ($n = 105$) to predict what the majority of subjects in the first group had indicated that one "should do" in the experiment (i.e., normative expectations), again choosing among three possible responses: take, give, or abstain. This second elicitation was incentivized: subjects were paid a monetary reward of appropriate size relative to the effort and duration of the experiment ($0.25) if their normative expectations matched the personal normative beliefs of the first group of subjects. (3) Finally, we asked a third group of subjects ($n = 100$), to report their empirical expectations. Subjects first read a description of the ToG game and were then told that a number of participants had previously made ToG choices in the game. They then had to predict which of the three options (give, take, abstain) the majority of those previous participants chose. Subjects were paid a small monetary reward ($0.25) if they predicted correctly, and nothing otherwise.

Second, as an additional approach to elicit normative expectations and confirm the robustness of our previous results, we followed the technique introduced in the economics literature by Krupka and Weber (2013). Subjects (N = 152, also recruited on Amazon Mechanical Turk) were given a description of the ToG donation game and were asked to rate the social appropriateness of various behaviors in the game. In the original procedure introduced by Krupka and Weber (2013), subjects are asked to rate each action available to the decision maker in the game. In our case, however, this would be too demanding since the game involves a very large action space. Consistent with the norm elicitation above, we therefore asked subjects to rate three types of behaviors: *take* money from the charity; *abstain* (leave initial endowments unchanged); or *give* money to the charity.

For each behavior, subjects were asked to report whether they perceived it as socially appropriate or socially inappropriate using a 4-point scale (very inappropriate; somewhat inappropriate; somewhat appropriate; very appropriate). Importantly, subjects were not asked to report their own perception about the appropriateness of the behavior, but rather their belief about the shared perception of appropriateness among individuals completing the same task as themselves. That is, subjects were asked to report their beliefs about what is commonly perceived as appropriate behavior in the game, rather than what they personally perceive as appropriate or inappropriate. We interpret this belief as measuring a subject's normative expectation about behavior in the game. In order to incentivize the elicitation of normative expectations, subjects were paid $0.25 if their rating of appropriateness for a behavior matched the modal response given by other participants.

III.b. Results (Experiment 3)

We present the results of both elicitations in Figures A.9 and A.10, respectively.

First, Figure A.9 presents three sets of beliefs. A subjects' personal normative beliefs are shown by the dark solid line, representing the percentages of subjects believing that one ought to take, give, or abstain from changing endowments in the ToG game. The figure shows that a majority of subjects believes that giving constitutes prescribed behavior (51% indicated so), while only 17% of subjects believe that one ought to take.
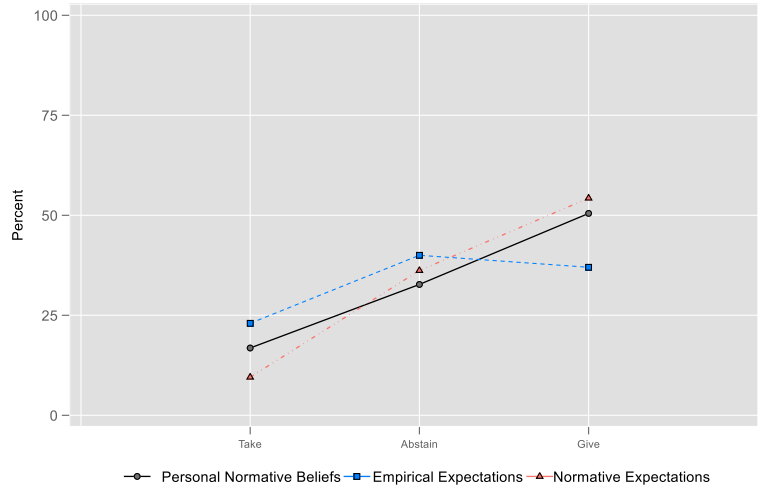

Figure A.9: Personal normative beliefs, normative expectations and empirical expectations in the ToG game using the Bicchieri and Chavez (2010) method.

26

The dashed red line in Figure A.9 shows subjects' normative expectations (second-order beliefs about what ought to be done in the game). The line shows that second-order beliefs track closely the corresponding (first-order) personal beliefs of prescribed behavior: 54% of subjects believe that a majority of others indicated that giving is what one ought to do in the game, and only 10% believe that the majority of others thought that one ought to take. A $\tilde{\chi}^2$ test shows that the distributions of personal normative beliefs and normative expectations are not significantly different from one another ($\tilde{\chi}^2(2) = 2.471$, p = 0.290). Together, the personal normative beliefs and normative expectations indicate that there are mutually consistent expectations among a majority of subjects that giving constitutes prescribed behavior in the experiment.

Finally, the dashed blue line in Figure A.9 represent subjects' empirical expectations. We find that 37% of subjects expect a majority of participants in the game to give to charity, and another 40% to leave initial endowments unchanged. Only 23% of subjects expect others to take in the game. Thus, although normative expectations clearly indicate that giving is what one should do, subjects' initial empirical expectations indicate that subjects expect most others to abstain from taking, rather than to actively give to charity. More generally, empirical expectations appear more heterogeneous than normative expectations and personal normative beliefs, with a sizeable fraction of subjects anticipating others to actually take from charity; a $\tilde{\chi}^2$ test shows significant differences between the distributions of empirical and normative expectations ($\tilde{\chi}^2(2) = 9.311$, p = 0.010). This is somewhat different from what we found in our sample of university students, where empirical expectations followed more closely normative expectations and personal beliefs, and these beliefs were more strongly supportive of giving relative to those elicited in Experiment 3. These differences are likely to be due to subject pool idiosyncrasies (students vs. MTurkers) and the fact that MTurkers seemingly have gloomier expectations of other MTurkers' behavior towards a charity compared to their student counterparts. Still, the results show that, overall, in both subject pools there is a clear norm against taking from charity. The pro-giving norm is stronger in our subject pool of lab students than in the pool of MTurkers.

Next, Figure A.10 illustrates the results from the Krupka and Weber (2013) elicitation. In the top panel, we plot, for each behavior, the average appropriateness rating given by subjects in the experiment (where a rating of -1 means "very inappropriate" and +1 means "very appropriate"). In the bottom panel, we plot the distribution of appropriateness ratings sub-divided by every possible action subjects evaluated in the experiment. Compared with our results in Figure A.7, Figure A.10 shows that, regardless of the method used

to elicit normative expectations, one arrives at a similar conclusion about the prevailing normative expectations in the ToG donation game. In both cases, giving money to the charity is rated as the most appropriate behavior or what one ought to do, while taking money away from the charity is the least appropriate behavior.
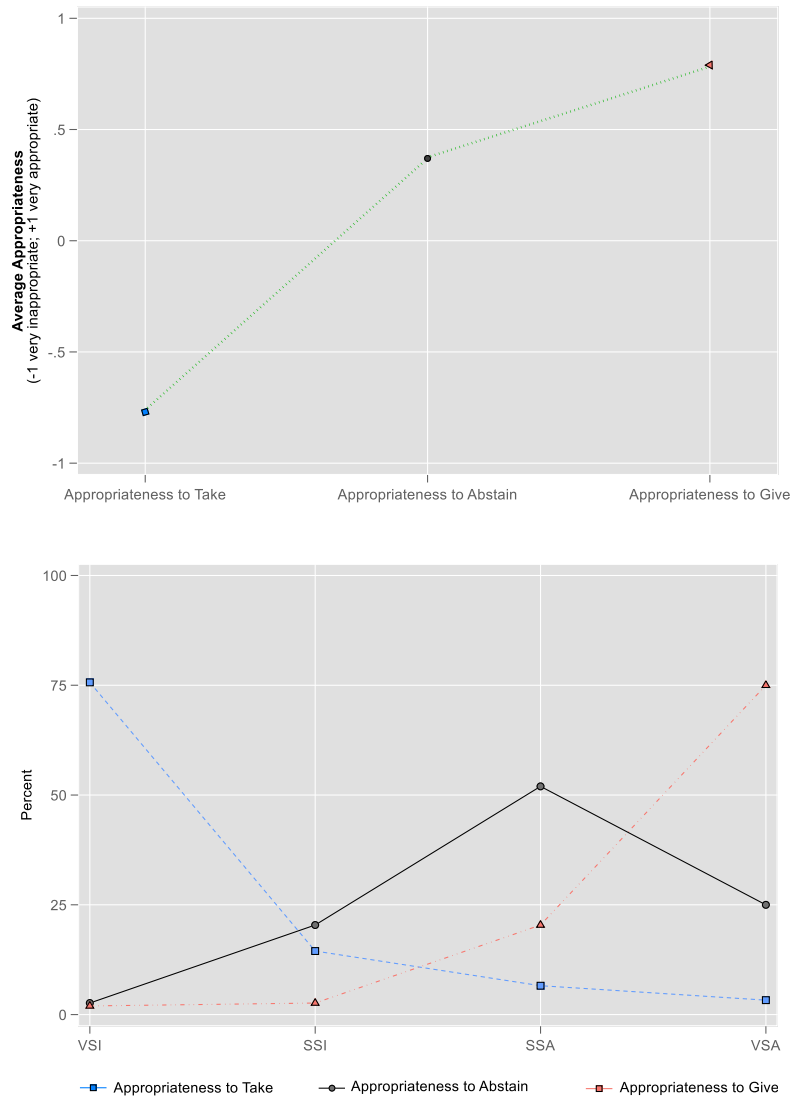

Figure A.10: Normative expectations elicited using the Krupka and Weber (2013) method. VSI = Very Socially Inappropriate; SSI = Somewhat Socially Inappropriate; SSA = Somewhat Socially Appropriate; VSA = Very Socially Appropriate. Top panel: average appropriateness ratings. Bottom panel: distribution of appropriateness ratings sub-divided by every possible action.

III.c. Experimental Screenshots (Experiment 3)

Below, we present the screenshots of the norm elicitation experiments (all between-design) that we ran on MTurk. As indicated, the first three and the last 2 screens were the same across all elicitations. The other screens correspond to the respective norm elicitations.

Consent Form (same for all elicitations)

University of Pennsylvania
Department of Philosophy, Politics, and Economics
Claudia Cohen Hall, Room 311
Philadelphia, PA 19104
Phone: (215)-898-3023
Fax: (215) 573-2231

Informed Consent/ Assent Form for Non-Pool Participants Earning Money

You are invited to take part in a study named *Guessing Task*. The purpose of this research study is to explore human decision-making. You will complete a series of computer tasks, each involving semantic as well as visual stimuli materials. If you agree to be in this study, you will need to make decisions and answer questions regarding the study materials. We will also ask you to provide demographic information. We will not ask for your name or any information that will make you identifiable. Overall, this study will take approximately 5-10 minutes.

For your participation in this study, you will receive a fixed payment of $0.25. Additionally, you may receive a monetary bonus. The exact amount depends on your results in the experiment. The risks to participating are no greater than those encountered in everyday life. Your participation in this study is completely voluntary, and you may refuse to participate or withdraw from the study without penalty or loss of benefits to which you may otherwise be entitled. Compensation will be awarded upon completion of the entire study.

Results may include summary data, but you will never be identified. If you have any questions about this study, you may contact the Behavioral Ethics Lab at behavioralethicslab@gmail.com.

For any questions, concerns, suggestions, or complaints that are not being addressed by the researcher, please contact the Institutional Review Board at the University of Pennsylvania, 3624 Market Street, Suite 301 South Philadelphia, PA 19104-6006. Phone: (215) 898-2614.

Please feel free to print or save a copy of this consent form.

By continuing from this page you are indicating that you have read and understood this consent form and wish to continue your participation in this study.

Consent

○

→

General Instructions (same for all elicitations)

Thank you for choosing to participate in this study! It is important that you <u>read all of the instructions carefully to maximize your earnings</u>. This is a survey-based study where you will be asked to answer questions and complete simple tasks. At the end of the study, you will be asked to fill out a short questionnaire and demographic questions.

It is expected that this study will take approximately 5-10 minutes. There is no deception in this study. Everything you see or read is true.

You begin the survey with an endowment of $0.25 (your fixed payment) and will have the opportunity to earn more based on your answers. You will be presented with a set of statements and asked to determine which actions you would be more likely to take after reading these statements. You will also be asked to guess what other participants have done previously. Guessing correctly will earn you an additional $0.25 each.

You may stop participating in this study at any time; however, you will not receive your payment unless you complete the entire survey through to the end. You will receive a completion at the end of the survey.

→

29

Norm Elicitation 1 Following Bicchieri & Chavez (2010)

Scenario

In what follows, we describe some scenarios to you based on a series of prior actual experiments. **While you will be making hypothetical decisions, the scenarios below describe the original experimental setup truthfully.** In particular, this experiment involved real stakes, and the taking or donation decisions towards real charities were implemented and paid accordingly to both the participants and the respective charities.

Please read the subsequent statements carefully and answer the questions accordingly. Carefully reading the question will help you to make additional earning in this survey.

In this setup, you and a charity (either Doctors Without Borders, World Wildlife Fund (WWF), or UNICEF) start with an equal endowment of $10 each (meaning that both you and the charity own $10 at the beginning of the experiment).

You are in the role of a **decision maker**. As a decision maker, you would have to decide whether to...

- ...**take** a part or all of the money **from the charity** and add this amount to the decision maker's account.
- ...**leave the equal division** of the sum of money as it is.
- ...**give** a part or all of the decision maker's money **to the charity** and add this amount to the charity's account.

Three Belief Elicitations (Between-Design)

1. *Personal Normative Beliefs (PNB) - Decision*

Based on the scenario described above, please tell us what you think one should do?

○ **Leave the equal division** of the sum of money as it is.

○ **Give** a part or all of the money **to the charity**.

○ **Take** a part or all of the money **from the charity** and add this amount to the decision maker's account.

→

2. Normative Expectations (NE) - Decision

Please guess what you believe other participants who participated in the survey **said**. If you guess correctly what other participants said, you will receive an additional $0.25.

Guess the correct statement: **The majority of participants in the survey said that one should...**

- ◉ Leave the equal division of the sum of money as it is.
- ○ Take a part or all of the money from the charity.
- ○ Give a part or all of the money to the charity.

→

3. Empirical Expectations (EE) – Decision

Please guess what you believe other participants who already participated in the experiment described above **did**. If you guess correctly what other participants who already participated did, you will receive an additional $0.25.

Guess: **Based on the behavior of participants who previously played the game in the role of a decision maker with the charity, the...**

- ○ Majority of participants left the equal division of the sum of money as it was.
- ○ Majority of participants took a part or all of the money from the charity.
- ○ Majority of participants gave a part or all the money to the charity.

→

Norm Elicitation 2 Following Krupka & Weber (2013)

In what follows, we describe some scenarios to you based on a series of actual prior experiments. The scenarios below describe the original experimental setup truthfully. In particular, this experiment involved real stakes and the taking or donation decisions towards real charities were implemented and paid accordingly to both the participants and the respective charities.

Please read the subsequent statements carefully and answer the questions accordingly. Carefully reading the question will help you to make additional money in this survey.

In this setup, you and a charity (either Doctors Without Borders, World Wildlife Fund (WWF), or UNICEF) start with an equal endowment of $10 each (meaning that both you and the charity own $10 at the beginning of the experiment).

You are in the role of a decision maker. As a decision maker, you would have to decide whether to...
- ...take a part or all of the money from the charity and add this amount to the decision maker's account.
- ...leave the equal division of the sum of money as it is.
- ...give a part or all of the decision maker's money to the charity and add this amount to the charity's account.

In what follows, you need to make payoff-relevant guesses.

Note: this survey will also be given to a number of other participants.
In order to earn additional money, you need to identify the answer that will be given the most often by the other participants. If you identify the most frequently given answers to the following statements, you will earn an additional $0.25 each per correct statement.

Please rate the appropriateness of the following behavior based on what you believe the most frequent answer will be in this survey:

	Very Socially Inappropriate	Somewhat Socially Inappropriate	Somewhat Socially Appropriate	Very Socially Appropriate
Taking a part or all of the money from the charity	○	○	○	○
Leaving the equal division of the sum of money as it is	○	○	○	○
Giving a part or all of the money to the charity	○	○	○	○

→

Post-Experimental Questionnaire (same for all elicitations)

What is your gender?

- Female
- Male

What is the highest level of school you have completed or the highest degree you have received?

- Less than high school degree
- High school graduate (high school diploma or equivalent including GED)
- Some college but no degree
- Associate degree in college (2-year)
- Bachelor's degree in college (4-year)
- Master's degree
- Doctoral degree
- Professional degree (JD, MD)

How old are you in years?

How do you see yourself:

Please indicate on the scale if are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?

0 means: `not at all willing to take risks'
10 means: `very willing to take risks'

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Risk Willingness

Below are a series of statements about you. Please read each statement and decide how much you agree or disagree with that statement.

Please answer every statement, even if you are not completely sure of your response.

If I knew that I could never get caught, I would be willing to steal a million dollars.	▼
I would never accept a bribe, even if it were very large.	▼
I'd be tempted to use counterfeit money, if I were sure I could get away with it.	▼
Having a lot of money is not especially important to me.	▼
I would get a lot of pleasure from owning expensive luxury goods.	▼
I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.	▼
If I want something from someone, I will laugh at that person's worst jokes.	▼
I wouldn't pretend to like someone just to get that person to do favors for me.	▼

→

What is your general opinion of charitable organizations?

'1' = I do not support the work of charitable organizations.
'10' = I fully support the work of charitable organizations.

1	2	3	4	5	6	7	8	9	10

General Opinion of Charities

What is your general opinion of the United Nations International Children's Emergency Fund (UNICEF)?

'1' = I do not support the work of UNICEF.
'10' = I fully support the work of UNICEF.

1	2	3	4	5	6	7	8	9	10

General Opinion of the UNICEF

What is your general opinion of the Doctors Without Borders?

'1' = I do not support the work of Doctors Without Borders.
'10' = I fully support the work of Doctors Without Borders.

1	2	3	4	5	6	7	8	9	10

General Opinion of the Doctors Without Borders

What is your general opinion of the World Wildlife Fund (WWF)?

'1' = I do not support the work of the WWF.
'10' = I fully support the work of the WWF.

1	2	3	4	5	6	7	8	9	10

General Opinion of the WWF

So that individual payments can be made, please enter your Amazon Mechanical Turk Worker ID. (This number can be found on your worker Dashboard)

Type carefully. If your Worker ID is entered incorrectly, additional earnings may not be received.

Worker ID:

After payments are made, your Worker ID will be permanently deleted from the experimenter's data.

Optional:
If you have any comments please feel free to write them here, thank you!

→

IV.a. Design (Experiment 4)

To study the mechanisms underlying the erosion of norm compliance we conducted one further experiment. The experiment was conducted with $N = 174$ students at the University of Pennsylvania who had not participated in any of our previous experiments. Participants received $2 upfront for participating and up to $4 if they answered the normative expectation questions correctly (see discussion below). The average duration of the experiment was 10 minutes. The procedures of this "expectations-updating" experiment were similar to those of the norm-elicitation experiment described in Appendix II. To elicit personal beliefs, participants were asked to report the action in the ToG game that they personally thought one should take. To elicit normative expectations, they were then incentivized to correctly guess the most common answer to that question. Participants were also incentivized to guess the most common action actually taken in the game. After this first round of elicitations, subjects were shown the actual behavior of participants in the main behavioral experiment using a figure similar to Figure 2 (left panel). Participants were randomly shown behavior in only one of the three treatments of the behavioral experiment (between-subject design). We only showed the dynamics of norm compliance without giving them any information about the nature of the treatments, so as to examine the impact of observing behavior on belief change without conflating it with any additional signals about the underlying data generating process. Thus, the only difference across the three conditions is the severity of the violations of the norm of giving (milder in NoObservation and ObservationSP than in Observation). After being shown the behavior in the main experiment (and hence having had to revise their initial empirical expectations accordingly), subjects had to report their personal beliefs and their (incentivized) normative expectations a second time.

IV.b. Results (Experiment 4)

We present our results in Figure A.11 below in the form of differences between the second belief elicitation and the first belief elicitation (i.e., negative values indicate that the second beliefs, after behavior was observed, were more negative than the initial beliefs).
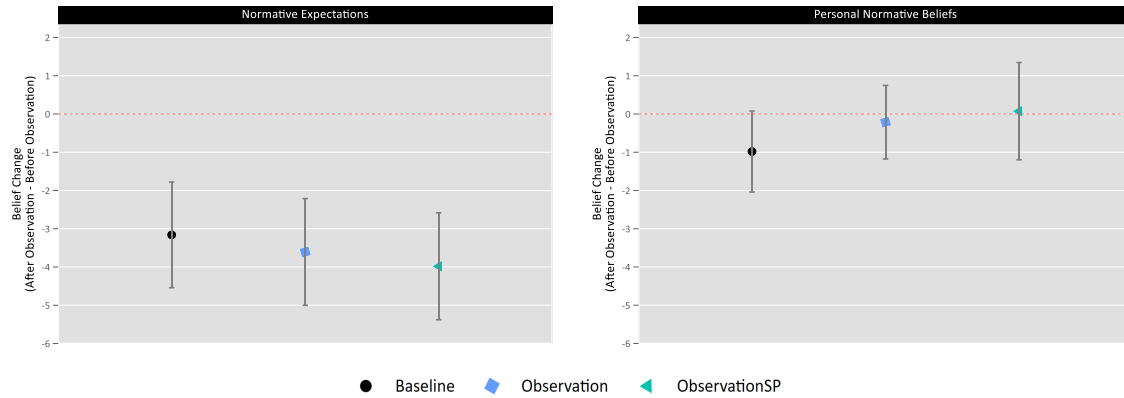

Figure A.11: Left panel: Changes in normative expectations. Right panel: Changes in personal normative beliefs across treatments. Negative (positive) values indicate that the beliefs after the observation of behavior were lower (higher) than beliefs prior to observation of behavior. Whiskers represent 95% CIs.

We find evidence of a significant indirect effect of observing behavior on normative expectations. In all cases, normative expectations (left panel Figure A.11) were adjusted significantly downwards after observing evidence of violations of the norm of giving (all p-values < 0.01). For personal beliefs (right panel Figure A.11), we observe small and insignificant revisions of beliefs. Overall, these results suggest that observing others' behavior has an indirect effect on norm compliance, as observation leads subjects to mainly update only their normative expectations.

IV.c. Experimental Screenshots (Experiment 4)

Consent Form

Informed Consent Form

You are invited to take part in a study named *Guessing Task*. The purpose of this research study is to explore human decision-making. You will complete a series of computer tasks, each involving semantic as well as visual stimuli materials. If you agree to be in this study, you will need to make decisions and answer questions regarding the study materials. We will also ask you to provide demographic information. We will not ask for your name or any information that will make you identifiable. Overall, this study will take approximately 5-10 minutes.

For your participation in this study, you will receive a fixed payment of $2. Additionally, you may receive a monetary bonus. The exact amount depends on your results in the experiment. The risks to participating are no greater than those encountered in everyday life. Your participation in this study is completely voluntary, and you may refuse to participate or withdraw from the study without penalty or loss of benefits to which you may otherwise be entitled. Compensation will be awarded upon completion of the entire study.

Results may include summary data, but you will never be identified. If you have any questions about this study, you may contact the Behavioral Ethics Lab at arjunkh@sas.upenn.edu or behavioralethicslab@gmail.com.

Please feel free to print or save a copy of this consent form.

By continuing from this page you are indicating that you have read and understood this consent form and wish to continue your participation in this study.

Consent
○

General Instructions

Thank you for choosing to participate in this study! It is important that you **read all of the instructions carefully to maximize your earnings**. This is a survey-based study where you will be asked to answer questions and complete simple tasks. At the end of the study, you will be asked to fill out a short questionnaire and demographic questions.

It is expected that this study will take approximately 5-10 minutes. There is **no deception** in this study. Everything you see or read is true.

You begin the survey with an endowment of $2 (your fixed payment) and will have the opportunity to earn more based on your answers. You will be presented with a set of statements and asked to determine which actions you would be more likely to take after reading these statements. You will also be asked to guess what other participants have done previously. Guessing correctly will earn you an additional $1 each.

You may stop participating in this study at any time; however, you will not receive your payment unless you complete the entire survey through to the end. You will receive a completion at the end of the survey.

Scenario

In what follows, we describe a scenario based on a previous experiment. The experiment involved real stakes and participants' decisions were actually implemented.

In the experiment, the participant and a charity started with an endowment of $10 each. The charity was chosen by the participant from the following three: Doctors without Borders, World World Wildlife Fund (WWF), and UNICEF.

The participant had to decide whether to:

- ...**take** a part or all of the money **from the charity**
- ...**make no change** to either of the endowments
- ...**give** a part or all of the their money **to the charity**

For example, making no change means choosing 0 on the following slider. Giving to a charity means choosing a positive amount on the slider. Taking from the charity means choosing a negative amount on the slider.

Taking									No change									Giving	
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10

Amount ($)

Below, we also give 3 examples showing people taking, giving, and making no change to the charity's endowment.

Taking									No change									Giving	
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10

Example: Taking $5 from the charity

Example: Giving $2 to the charity

Example: Making no change to the charity

Three Belief Elicitations (Within-Design) – Before Observation Stage

1. Personal Normative Beliefs (PNB) - Decision

Based on the scenario described above, please mark on the slider **what you think one should do**

Taking   No change   Giving
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10

Amount ($)

How confident are you about your choice?

No confidence	Slight confidence	Moderate confidence	High confidence
○	○	○	○

2. Normative Expectations (NE) - Decision

Now, you will make a payoff relevant guess. You will earn an additional $1 if your guess is correct or if it is one unit above or below the correct answer.

This experiment has already been played. We have data about what participants in this experiment did.

Please guess what you believe the **most frequent choice the participants made** in the experiment about taking, giving, or making no change to the charity.

Taking   No change   Giving
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10

Amount ($)

How confident are you about your choice?

No confidence	Slight confidence	Moderate confidence	High confidence
○	○	○	○

3. Empirical Expectations (EE) – Decision

Now, you will make a payoff relevant guess. You will earn an additional $1 if your guess is correct or if it is one unit above or below the correct answer.

We have asked all participants in this survey what **they believe should be done**. Please guess what you believe is the **most frequent answer** that the participants in this survey give about what they believe should be done.

Taking   No change   Giving
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10

Amount ($)

How confident are you about your choice?

No confidence	Slight confidence	Moderate confidence	High confidence
○	○	○	○

Explanation Observation of Behavior: One of the three conditions (NoObservation, Observation, ObservationSP) is presented at random (Between-Design)

In what follows, we show you **real participant behavior** from a previous experiment. Those participants played the game (as it was described above) over several periods with a charity.

As a reminder:

- At the beginning of each round, both the participant and the charity started with an endowment of $10.
- The participant then had the opportunity to either leave the equal split as is, take some (or all) of the money from the charity and add it to one's own account, or give some (or all) of one's own money and add it to the charity account.

The illustration below can be read as follows:

- **x-axis (horizontal):** the period of play
- **y-axis (vertical):** change in the charity box (in % of the $10 that the charity started with)
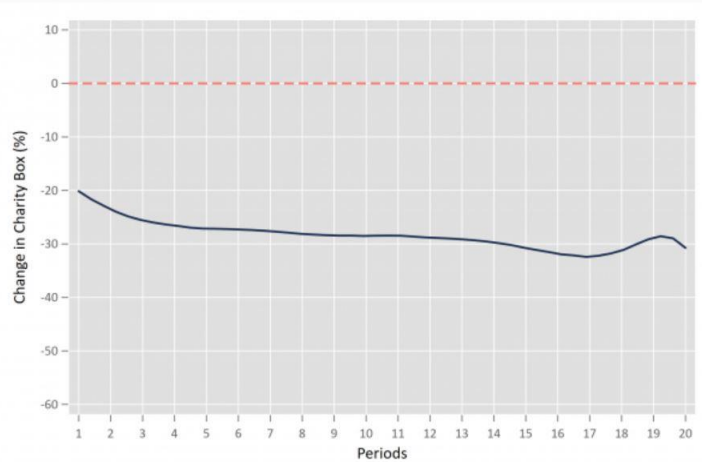
You will observe the *average behavior* of all participants towards the charity box in a particular condition.

- A **positive value** indicates that **more money was given** to the charity **than taken** from the charity
- Conversely, a **negative value** indicates that **more money was taken** from the charity **than given** to the charity

○ Understood

41

Presented Behavior: NoObservation (Baseline)

Below is the behavior (blue line) of previous participants towards their charity in the original experiment over the course of 20 rounds. Values below (above) zero indicate that participants took money away from (gave money to) charity. A value of 0 indicates that an equal distribution (= initial starting point) of money was maintained.

○ Understood

42

Below is the behavior (blue line) of previous participants towards their charity in the original experiment over the course of 20 rounds. Values below (above) zero indicate that participants took money away from (gave money to) charity. A value of 0 indicates that an equal distribution (= initial starting point) of money was maintained.
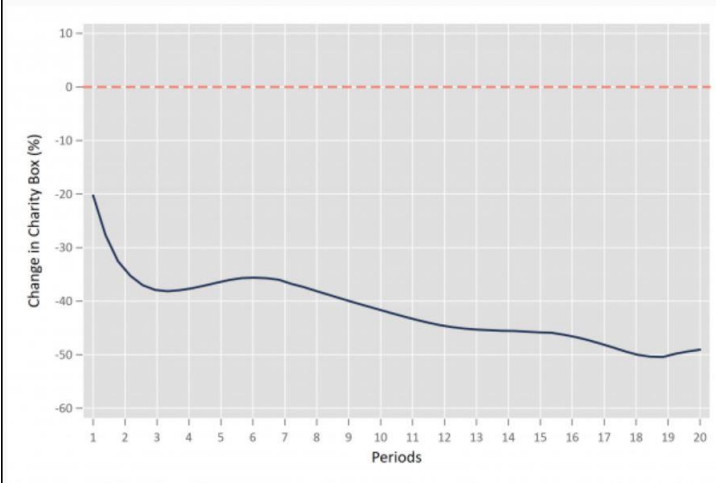
O Understood

Two Belief Elicitations (Within-Design) – After Observation Stage

1. Personal Normative Beliefs (PNB) - Decision

Now that you have observed the behavior from the previous experiment:

please mark on the slider **what you think one should do**

Taking										No change									Giving
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9 10

Amount ($)

How confident are you about your choice?

No confidence	Slight confidence	Moderate confidence	High confidence
○	○	○	○

2. Normative Expectations (NE) - Decision

You will make again a payoff relevant guess. You will earn an additional $1 if your guess is correct or if it is one unit above or below the correct answer.

Now that you have observed the behavior from the previous experiment:

We have asked all participants in this survey again (after they also saw the behavior from the experiment) what **they believe should be done**. Please guess what you believe is the **most frequent answer** that the participants in this survey give about what they believe should be done.

Taking										No change									Giving
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9 10

Amount ($)

How confident are you about your choice?

No confidence	Slight confidence	Moderate confidence	High confidence
○	○	○	○

44

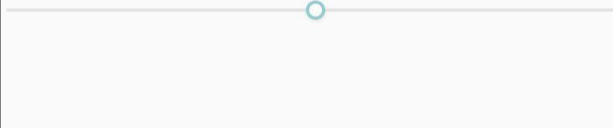**Two Belief Elicitations (Within-Design) – After Observation Stage**

1. Personal Normative Beliefs (PNB) - Decision

Now that you have observed the behavior from the previous experiment:

please mark on the slider **what you think one should do**

Taking										No change									Giving
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9 10

Amount ($)

How confident are you about your choice?

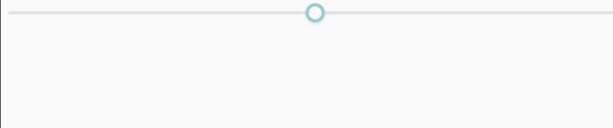| No confidence	Slight confidence	Moderate confidence	High confidence
○	○	○	○

2. Normative Expectations (NE) - Decision

You will make again a payoff relevant guess. You will earn an additional $0.25 if your guess is correct or if it is one unit above or below the correct answer.

Now that you have observed the behavior from the previous experiment:

We have asked all participants in this survey again (after they also saw the behavior from the experiment) what **they believe should be done**. Please guess what you believe is the **most frequent answer** that the participants in this survey give about what they believe should be done.

Taking										No change									Giving
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9 10

Amount ($)

How confident are you about your choice?

No confidence	Slight confidence	Moderate confidence	High confidence
○	○	○	○

45

Post-Experimental Questionnaire

What is your gender?

○ Female

○ Male

What is the highest level of school you have completed or the highest degree you have received?

○ Less than high school degree

○ High school graduate (high school diploma or equivalent including GED)

○ Some college but no degree

○ Associate degree in college (2-year)

○ Bachelor's degree in college (4-year)

○ Master's degree

○ Doctoral degree

○ Professional degree (JD, MD)

How old are you in years?

How do you see yourself:

Please indicate on the scale if are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?

0 means: `not at all willing to take risks'
10 means: `very willing to take risks'

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Risk Willingness

○————————————————————————

What is your general opinion of charitable organizations?

'1' = I do not support the work of charitable organizations.
'10' = I fully support the work of charitable organizations.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

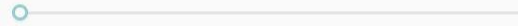General Opinion of Charities

○————————————————————————

What is your general opinion of the United Nations International Children's Emergency Fund (UNICEF)?

'1' = I do not support the work of UNICEF.
'10' = I fully support the work of UNICEF.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

General Opinion of the UNICEF

○————————————————————————

What is your general opinion of the Doctors Without Borders?

'1' = I do not support the work of Doctors Without Borders.
'10' = I fully support the work of Doctors Without Borders.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

General Opinion of the Doctors Without Borders

○————————————————————————

What is your general opinion of the World Wildlife Fund (WWF)?

'1' = I do not support the work of the WWF.
'10' = I fully support the work of the WWF.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

General Opinion of the WWF

○————————————————————————