CESifo WORKING PAPERS

8212
2020

April 2020

Observability, Social Proximity, and the Erosion of Norm Compliance

Cristina Bicchieri, Eugen Dimant, Simon Gächter, Daniele Nosenzo

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8212*

Observability, Social Proximity, and the Erosion of Norm Compliance

Abstract

We study how an individual‘s compliance with social norms is inuenced by other actors‘ norm compliance. In a repeated non-strategic Take-or-Give donation experiment we show that giving is considered socially appropriate while taking is socially inappropriate. Observing norm violations erodes an individual‘s own norm compliance. We show that the erosion of norm compliance is led by a change in norm-related beliefs. This change has a major effect on individuals who initially obey the norm, driving them to non-compliance, whereas initially non-compliant individuals do not change their taking behavior. Erosion is halted when individuals have even minimal social proximity to those they observe: individuals now also pay attention to norm followers.

JEL-Codes: C920, D640, D900.

Keywords: norm compliance, social norms, social proximity.

Cristina Bicchieri
University of Pennsylvania, Center for Social Norms and Behavioral Dynamics
Philadelphia / PA / USA
cb36@sas.upenn.edu

*Eugen Dimant**
University of Pennsylvania, Center for Social Norms and Behavioral Dynamics
Philadelphia / PA / USA
edimant@sas.upenn.edu

Simon Gächter
University of Nottingham
United Kingdom - NG7 2RD, Nottingham
Simon.Gaechter@nottingham.ac.uk

Daniele Nosenzo
University of Nottingham
United Kingdom - NG7 2RD, Nottingham
Daniele.Nosenzo@nottingham.ac.uk

*corresponding author

This version: March 24, 2020
We would like to thank Johannes Abeler, Damon Centola, Ernst Fehr, Stefano Fiorin, Bernd Irlenbusch, Erin Krupka, George Loewenstein, Michel Maréchal, Nikos Nikiforakis, Dirk Sliwka, Bertil Tungodden, and Elke Weber for helpful comments and suggestions. We also appreciate discussions with attendees at the 2018 Cultural Evolution Society Conference, 2019 CESifo Area Conference on Behavioral Economics, 2019 Science of Philanthropy Initiative (SPI) Conference at the Unversity of Chicago, the 3rd Cultural Transmission and Social NormsWorkshop at MIT, as well as seminar participants at the Behavioral Insights Team UK, the Choice Lab at Bergen, London School of Economics, NYU Abu Dhabi, Stockholm School of Economics, Wharton Business School, and the Universities of Amsterdam, Cologne, Lyon, Oxford, Princeton, and Yale. This work was supported by the European Research Council [grant number ERC-AdG 295707 COOPERATION] and the Economic and Social Research Council [grant number ES/K002201/1]. The most recent version of the working paper can always be downloaded following this link: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3355028

1. Introduction

Social norms guide our behavior and interactions in a variety of economically interesting domains. They constrain anti-social acts like cheating, corruption and tax evasion (Fisman and Miguel, 2007; Fellner et al., 2013; Hallsworth et al., 2017; Kocher et al., 2017; Muthukrishna et al., 2017; Bicchieri et al., 2019a; Bott et al., 2019), abuse of the welfare state (Lindbeck et al., 1999), youth delinquency (Kling et al., 2005; Chyn, 2018), and discrimination (Barr et al., 2018). Social norms also support pro-social behaviors, such as collective actions (Ostrom, 2000), environmental conservation (Goldstein et al., 2008), voter turnout (Knack, 1992; Gerber et al., 2008), fairness in negotiations (Bartling and Schmidt, 2015), charitable giving, and altruistic sharing (DellaVigna et al., 2012; Krupka and Weber, 2013; Krupka and Croson, 2016; Gächter et al., 2017; Bašić et al., 2019; Gächter et al., 2013; Kimbrough and Vostroknutov, 2016; Bicchieri et al., 2019b).

At the same time, despite progress in uncovering the importance of social norms in many economically interesting settings, our understanding of the factors that influence individuals' willingness to *comply* with social norms is still incomplete (for a recent discussion see Bicchieri, 2016; Bicchieri and Dimant, 2019). This is important because norm compliance can quickly erode in response to changes in the social, physical, or political landscape, with detrimental effects whenever norms serve the purpose of curbing anti-social or promoting pro-social behavior (Goldin and Katz, 2002; Keizer et al., 2008; Andreoni et al., 2017; Bartling and Özdemir, 2017; Centola, 2018; Bursztyn et al., 2019; Falk et al., 2020).

In this paper, we focus on two potential factors that may influence the willingness to comply with norms: observations of other people's behavior and the degree of social proximity individuals have with those whose behavior they observe. Our study is guided by the conceptual framework of social norms based on preferences and beliefs developed by Bicchieri (2006, 2016) and relies on a novel experimental design conducted in a controlled laboratory environment. Expanding on the tradition of List (2007) and Bardsley (2008), we introduce a novel repeated Take-or-Give (ToG) donation game in a *non-strategic* group setting in which decision-makers can either give money to a charity, take money from it, or retain the initial equal split. We ran several experiments to answer our research question exhaustively, including one that establishes the existence of a social norm in our setting, one that examines norm compliance change in the form of change of normative expectations, and one that tracks behavior and norm compliance (or the lack thereof)

in the ToG donation game over time, under three different information conditions that systematically vary observability of behavior and social proximity. The ToG game allows us to study both norm-following (giving money to a charity) and norm-breaking (taking money from charity) in the absence of strategic elements as present in games such as Public Goods or Ultimatum Games (Bohnet and Zeckhauser, 2004; Neugebauer et al., 2009). We established the *existence* of a norm in our setting by applying the measures introduced by Bicchieri and Chavez (2010) to assess subjects' personal normative beliefs, empirical expectations, and normative expectations about donating to charity. All measures provide consistent evidence that players consider charitable giving to be a social norm and taking to be a norm violation.

Next, we study *behavior* in the ToG donation game with a separate group of participants. Subjects make repeated and independent decisions in this game across 20 periods. Between treatments, we systematically vary whether players 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 receive no feedback about others' ToG decision. In the 'Observation' treatment, players are informed after each round about the ToG decisions of their group members. In our final treatment, 'ObservationSP', we combine Observation with social proximity. Group members can observe others' ToG decisions after each round but also receive information about their social proximity to the other group members on one particular and fairly minimal dimension that does not permit to further identify players (we used fandom for a local sports team). Note that our design implements the decisions of others in a *non-strategic* and *anonymous* way in all our treatments, so that there is no monetary or strategic impact from learning others' behavior.

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. 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. Knowledge of social proximity strongly moderates ToG donation behavior. In ObservationSP, the average donation behavior is not significantly different from our NoObservation baseline. The reason is that individuals respond to both observed norm compliances and violations in socially

proximate groups. In the presence of known social proximity, both forms of behavior are contagious and stabilize donations roughly at their initial level.

Finally, we show that there is heterogeneity in the extent to which subjects respond to feedback about others' behavior. The biggest changes in behavior come from subjects who are norm-compliant and give to charity before being exposed to others' negative behavior. In contrast, subjects who initially take from charity, and thus violate the norm, do not respond to positive or negative feedback about others' behavior. The response from subjects who initially abstain falls between these two groups.

Our paper makes five contributions to a deeper understanding of social norm compliance. First, while most of the existing literature focuses on showing the existence of norms, we use a framework that enables us to distinguish between the *existence* of norms and the extent to which they are *followed*. Instead of using a definition of norms that relies only on normative expectations (beliefs about what others think is appropriate behavior) (Gächter et al., 2013; Krupka and Weber, 2013; Fehr and Schurtenberger, 2018), we focus on a richer definition that includes both empirical expectations (beliefs about others' behavior) and personal normative beliefs (personal beliefs about appropriate behavior) (Bicchieri, 2006, 2016). This definition allows us to better focus on the drivers and dynamics of norm compliance, where empirical expectations play a crucial role.

Second, we compare the dynamics of norm compliance in a setting where individuals receive continuous feedback on others' behavior (treatment Observation) to one in which no such feedback is provided (baseline NoObservation). Our results suggest that updates of empirical expectations, which we induced by randomly grouping participants and letting them observe what others actually do, play a key role in the dynamic of norm compliance. This result is reminiscent of behavior in a repeated public goods game, where contributions do not decline without feedback about others' contributions, but do decline with such feedback (Neugebauer et al., 2009). Crucially, our results reveal that observing others' behavior has a causal influence on the observer's behavior in a *non-strategic* setting.

Third, we show that in the Observation treatment the response to observed behavior is asymmetric: while individuals respond strongly to others' norm violations, the response to examples of norm compliance is weak. As a consequence, we observe a marked erosion of norm compliance over time when information about others' behavior is freely available. This finding extends recent work on peer effects in norm compliance, which also shows (in static, one-shot settings, in contrast with our dynamic, repeated setting) that individuals mainly respond to examples of norm violations, much less to examples of norm compliance

(e.g. Gino et al., 2009a; Dimant, 2019; Gächter et al., 2019).

Fourth, and most importantly, 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 (in ObservationSP), we observe strong responses not only to norm violations, but also to examples of norm compliance. This mitigates norm erosion in comparison to less socially proximate groups. This result highlights the importance of the broader social context in driving norm compliance. It also illustrates that punishment of norm violations, frequently seen as crucial to sustain norms (e.g., Coleman, 1994; Sutter et al., 2010; Fehr and Schurtenberger, 2018), is not always needed: even minimal social proximity can prevent the erosion of norm compliance.

Our fifth novel contribution to the existing literature is the ability to study whether there is heterogeneity in the erosion of norm compliance. Our design, explained in more detail below, allows us to identify subjects' initial dispositions to compliance prior to the observation of others' behavior. We use this design feature to differentiate between initial types of norm-followers (engaging in pro-social behavior towards a charity) and norm-violators (engaging in anti-social behavior towards a charity), and trace how their behavior changes when they are exposed to information about others' behavior. Our results indicate that initial norm-violators are largely insensitive to observation of peer behavior, whereas initial norm-followers are strongly influenced and their norm compliance erodes quickly.

In sum, we show that observation of peer behavior, and the social proximity to those peers, have a strong influence on the dynamics of norm compliance. Our results highlight the importance of empirical expectations for norm compliance. In a setting like ours, in which individuals recognize the existence of a social norm for giving, observing behavior that is in contrast with that norm (i.e. examples of taking) induces a (downward) updating of empirical expectations and thus weakens the normative pull (Bicchieri and Xiao, 2009). In comparison, examples of norm compliance have a weaker influence. There are two mechanisms behind this effect, one direct and the other indirect. The direct mechanism occurs because individuals' willingness to comply with norms depends on whether they believe (observe) that others are also complying. Our results point to the asymmetric nature of this conditional willingness to comply, possibly because examples of norm violations bolster selfish motives whereas examples of compliance require the individual to overcome their narrow self-interest. Moreover, we also show in a follow-up experiment that observing non-compliance has an indirect effect in that it leads to a (downward) revision of individuals' normative expectations about whether one should comply with the norm. Crucially, our

main result highlights that social proximity can mitigate the detrimental effects of being exposed to norm violations. Focusing individuals on social identity (even a minimal one) seems to induce sensitivity to *both* examples of compliance and violations, as individuals pay greater attention to and tend to emulate the behavior of those who are similar to them. This is an important finding, which shows that studying social norm compliance only in fully anonymous and socially sterile settings overestimates the erosion of norm compliance.

2. Social Norms: A Conceptual Framework

The concept of social norms has been used in various disciplines, from economics to sociology to psychology (e.g., Elster, 1989; Coleman, 1994; Ostrom, 2000; Cialdini and Goldstein, 2004; Burke and Young, 2011; Fehr and Gächter, 2000; Bicchieri, 2006; Fehr and Schurtenberger, 2018). The precise definition of what constitutes a social norm varies across (and within) these disciplines, although norms are often understood as behavioral patterns that are predominant within a group, supported by a shared understanding of acceptable actions, and sustained through that group's social interactions (for instance Ostrom, 2000). One issue with most existing definitions of social norms, including the one outlined above, is that they are difficult to operationalize in empirical research since it is unclear how to practically measure norms or assess their influence on behavior. Here, we follow the definition of social norms proposed by Bicchieri (2006). Unlike other approaches, this definition is firmly grounded within a belief and preference framework and is therefore easier to operationalize empirically (for a discussion see Bicchieri, 2016).

Our starting point is the notion that a norm is a special type of behavioral rule that prescribes or proscribes a certain behavior to a specific group of people in a specific class of situations. We say that the rule satisfies the properties of a social norm if the individual prefers to follow the rule on the condition that (a) she believes that sufficiently many others also follow it (empirical expectation) and (b) she believes that sufficiently many others believe that one should follow the rule and/or may be willing to sanction transgressions (normative expectation) (Bicchieri, 2006). Finally, we say that a norm exists within a group of people if the rule is known to apply within that group and subjects' normative expectations (i.e. what they believe others view as appropriate) and empirical expectations (i.e. what they believe others actually do) are *mutually consistent* (i.e. they all point to the same behavioral rule). Note that our approach to social norms ultimately defines norm compliance in terms of two underlying primitive constructs: beliefs about what is

commonly done and what is commonly approved of, and preferences to undertake a specific behavior that are conditional on such beliefs. A key advantage of defining a social norm in terms of preferences and beliefs is that we can independently measure and quantify these primitive constructs (and hence norms) using the methodology of experimental economics. Belief-elicitation protocols (e.g., for normative expectations as proposed by Bicchieri and Chavez, 2010, or Krupka and Weber, 2013) can be used to measure whether individuals hold sufficiently high and consistent empirical or normative expectations, and hence can be used to determine whether a majority believes that a norm applies to a given situation.

A second important advantage of our definition is that it makes it easier to differentiate between social behaviors that are regulated by social norms as opposed to behaviors regulated by other types of behavioral rules. A crucial distinguishing feature of behaviors that are regulated by social norms is that the motivation to undertake such behaviors is *conditional* on a person's belief about what is commonly done *and* what is commonly approved of within that person's reference network. The psychological foundation of such conditionality lies in the notion that humans naturally strive to obtain approval and avoid disapproval from others (e.g., Sugden, 1998, 2000; Brennan et al., 2004; Bicchieri, 2006, 2016). In contrast, the motivation to follow other types of behavioral rules may only be conditional on a person's belief about what is commonly done, as in the case of "descriptive norms" (Cialdini, 2007), or unconditional on her social expectations, as in the case of "moral rules" (Bicchieri, 2006). Assessing beliefs and preferences (and whether preferences depend on beliefs) lets us clearly distinguish between types of social behaviors, which is an important step in designing interventions to change behavior.

Finally, a crucial advantage of our approach is that it allows us to distinguish between the notion that a norm exists and the notion that it is followed. The existence of a norm depends on the mutual consistency of individuals' normative and empirical expectations. Since preference to comply with the norm is conditional on such expectations, a norm may exist without being followed at a particular time (Bicchieri, 2006) because expectations may change or weaken. If transgressions become common, or are no longer disapproved of, the norm will weaken and eventually disappear. For instance, consider two communities that hold similar normative expectations about the inappropriateness of littering: in both cases, individuals believe that littering is inappropriate, i.e. an anti-littering norm exists. In one community, this general social disapproval for littering is accompanied by evidence that littering is rare; empirical and normative expectations are both strong and aligned. In the other community, however, the disapproval of littering is accompanied by widespread

evidence that littering is becoming common. Empirical expectations of compliance are significantly weakened and observations are now incongruent with normative expectations, likely weakening them. If individuals in the latter community observe a sufficient number of transgressions, they may start littering, too, since their compliance is conditional upon what others actually do, not just what they say they *should* do. As a consequence, compliance (i.e. norm-following) may become much lower in the latter community, even if all members previously held similar beliefs about what is socially appropriate. Situational factors may also play an important role in determining whether or not a norm is followed; e.g., privacy can reduce compliance (Allen, 1965) and punishment can enhance it (Fehr and Gächter, 2000; Sutter et al., 2010; Fehr and Schurtenberger, 2018, but see also findings by Abbink et al. 2017 and Bicchieri et al. 2019b).

In our experiment, a group of anonymous individuals play a repeated non-strategic ToG donation game where they can take or give money from or to a charity across three treatments. We systematically vary whether subjects can observe the donation behavior of other members of their group in previous periods (NoObservation and Observation treatments, respectively), and the degree of known social proximity between group members (ObservationSP treatment). In a separate norm-elicitation experiment, we present direct evidence for the social norm that prescribes giving to and proscribes taking from the charity. We find that individuals' normative expectations are congruent with their own personal normative beliefs and their empirical expectations. This rules out the possibility that our observation is an instance of 'pluralistic ignorance', whereby people privately dislike a norm but believe that most others endorse and follow it (see Smerdon et al., 2019 for recent experimental evidence). With a separate group of individuals in our ToG donation game, we study how norm compliance is affected by feedback on what other group members have done in previous periods (by comparing Observation and NoObservation treatments). In our framework, feedback on the actions of others in previous periods constitutes information about what is commonly done in the experimental game. We expect this information to influence a subject's empirical expectations and hence their behavior, since norm-following partly depends on what others commonly do. Thus, we expect subjects to stop following the norm once they have enough evidence of others violating it. Should we similarly expect subjects to follow the norm after observing that others are also following it?

Although our conceptual framework does not make any explicit prediction about the symmetry or asymmetry of empirical information effects, the existing evidence suggests that these effects may be asymmetric: subjects strongly reduce compliance with norms when

they receive evidence that others are transgressing, but they only weakly increase compliance if they observe that others comply (Croson and Shang, 2008; Thöni and Gächter, 2015; Dimant, 2019; Gächter et al., 2019). One explanation for this asymmetry is that social norms involve a tension between selfish and other-regarding goals, as is the case for the norm of giving in our experiment. Since compliance with norms is costly from the point of view of individual material self-interest, individuals may look for ways to discount collective benefits and rationalize self-interested behavior. Thus, they may make a self-serving use of empirical information by responding more strongly when it is in their interest to do so (McDonald et al., 2013; Charness et al., 2019; Dimant et al., 2020). Consistent with this interpretation is evidence that individuals revise their normative expectations after receiving information about others' behavior. That is, the frequency of an observed behavior influences its acceptability status (Lindström et al., 2018; Bicchieri et al., 2019a).

The above discussion paints a rather pessimistic picture about the dynamics of norm compliance within groups that interact repeatedly and where individuals can observe each other's past behavior. If examples of norm violations are indeed more contagious than examples of compliance, we would expect to see an inexorable decline in norm-following over time in the absence of mechanisms that can effectively contrast the negative influence of bad examples. Are there means to reduce the erosion of norm compliance? Previous research has explored the effects of making normative expectations salient (Cialdini and Goldstein, 2004; Schultz et al., 2007; Jachimowicz et al., 2018). However, it has been shown that the combination of incongruent empirical and normative information is detrimental to norm compliance when the empirical information points to norm violation (Bicchieri and Xiao, 2009). One hitherto unstudied possibility is the role of *social identification* in moderating the effect of observed norm deviance. Social identification is defined as "the individual's knowledge that he/she belongs to certain social groups together with some emotional or value significance to him/her of the group membership" (Tajfel, 1982). Existing research has found that individuals care about the behavior of a group if they identify with the group, and perceived group identification depends upon recognizing similarities between the self and other group members (Perkins, 2002). According to social identity theory, identification with a group promotes behavioral conformity because when the individual self-categorizes as a group member, he/she seeks to undertake behaviors that are perceived as "what good group members do" and avoid behaviors that are perceived as inappropriate (Turner, 1985; Akerlof and Kranton, 2000; Chen and Li, 2009). These perceptions of appropriateness are informed by the behavior of other group members, and conforming to

perceived appropriate behavior signals commitment to the group (Hogg, 1992).

These considerations imply that an accentuation of similarities between self and other in-group members (what we call the "social proximity" between individuals) will lead the individual to give greater weight to empirical information about what other group members do, both because this information may be viewed as a direct source of what is perceived as "good group behavior" and because deviations from group behavior may signal low commitment to the group and thus trigger social sanctions by other in-group members. In terms of our conceptual framework, we can model these mechanisms as a heightened conditionality of preferences on empirical expectations in settings where individuals interact with socially proximate others. That is, individuals will attach greater significance (and respond more strongly) to the behavior of others when the others are in-group members than when they are anonymous strangers. Recent experimental evidence suggests that self-identification with the group drives conformity even when what is being followed is simply a descriptive norm, devalued of normative content (Pryor et al., 2019). In this case, greater weight assigned to in-group behavior occurs even in the absence of commitment signaling or the existence of sanctions. When behavior is driven by a social norm, we may expect the self-serving use of empirical information to be less prevalent among socially proximate individuals. Since non-conformity may be negatively judged by the group, it may be more costly for the individual to rationalize self-interested behavior by ignoring evidence of norm compliance when this evidence comes from the behavior of in-group members.

Thus, we expect a different dynamic of norm compliance to apply within groups that observe the past behavior of socially proximate others (ObservationSP treatment), as compared to groups that observe the past behavior of others who are not known to be socially proximate based on our measure (Observation treatments). In the Observation treatment, the existing evidence about asymmetric empirical information effects suggests the decline of norm compliance. In the ObservationSP treatment, we expect this process of erosion to be halted or substantially reduced, since individuals will respond to both examples of norm violation and norm compliance due to group identification effects.

3. Experiments

We conduct three experiments, for which we collected data from a total of $N = 1,126$ participants (plus $N = 303$ Amazon Mechanical Turk participants to test the robustness of some of our results). All experiments rely on a Take-or-Give (ToG) donation game

described in Section 3.1. In the first experiment (the "norm-elicitation" experiment, described in Section 3.2), we use the Bicchieri and Chavez (2010) norm-elicitation procedure to establish whether a social norm exists in the ToG game. We ask participants to describe their normative and empirical expectations, as well as their personal normative beliefs, in order to identify which norm, if any, applies to the setting.

In the second experiment (the "behavioral" experiment, described in Section 3.3), which we carried out with a different set of participants, we measure how participants behave in the ToG game. The design of this behavioral experiment consists of several treatments where we vary whether subjects can observe other participants' behavior as well as the degree of known social proximity between themselves and other participants. We can thus measure the extent to which participants comply with or violate the norm of giving that applies, and how norm compliance is affected by receiving information about the behavior of others and their social proximity.

Finally, in a third experiment (see detailed discussion in Section 4.2.4), we explore the mechanisms behind the behavioral effects of compliance and violations observations by relating them to how subjects update their personal normative beliefs and normative expectations. We call this experiment the "normative beliefs-updating" experiment.

3.1. The Take-or-Give Donation Game

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_1 is an integer representing the monetary amount which the i^{th} 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). A non-strategic setup is necessary to study the changes in norm compliance in absence of strategic elements that are present in related games such as public good games or ultimatum games (e.g., Bohnet and Zeckhauser, 2004).

3.2. Norm-Elicitation Experiment

We designed one separate experiment to elicit subjects' (1) personal, (2) normative, and (3) empirical expectations about behavior in the ToG game. As is the case for our main behavioral experiment (described below), this norm-elicitation experiment was conducted with students from the University of Pennsylvania, but only with those who have not previously participated in any of our other experiments.[1]

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 normative 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 normative belief question by other participants in the same experiment. These second-order beliefs represent subjects' *normative expectations* about what constitutes socially appropriate behavior in the game. 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 experiment was completed online with an average duration of 10 minutes. As we show in Section 4.1, these elicitations clearly indicate that a norm of giving to charity exists in the ToG game. Based on these results, we refer to giving to the charity as *norm-compliant* behavior and taking from the charity as *norm-violating* behavior.

3.3. Behavioral Experiment

To study how participants actually behave in the game, we conducted a behavioral experiment where we recruited $N = 842$ participants (University of Pennsylvania students)

[1] 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.

to make choices in the ToG donation game, none of whom have participated in the norm-elicitation experiment. Participants were on average 22 years old and 70 percent were female. The experiment was conducted in zTree (Fischbacher, 2007) and run in 52 sessions in the Behavioral Ethics Lab at the University of Pennsylvania across three between-subject treatments, described in more detail below. We have three treatments across which we vary the amount of information that participants receive in each period of Part III about the ToG donation choices of other members of their group. The four parts as illustrated in Figure 1 were introduced sequentially.

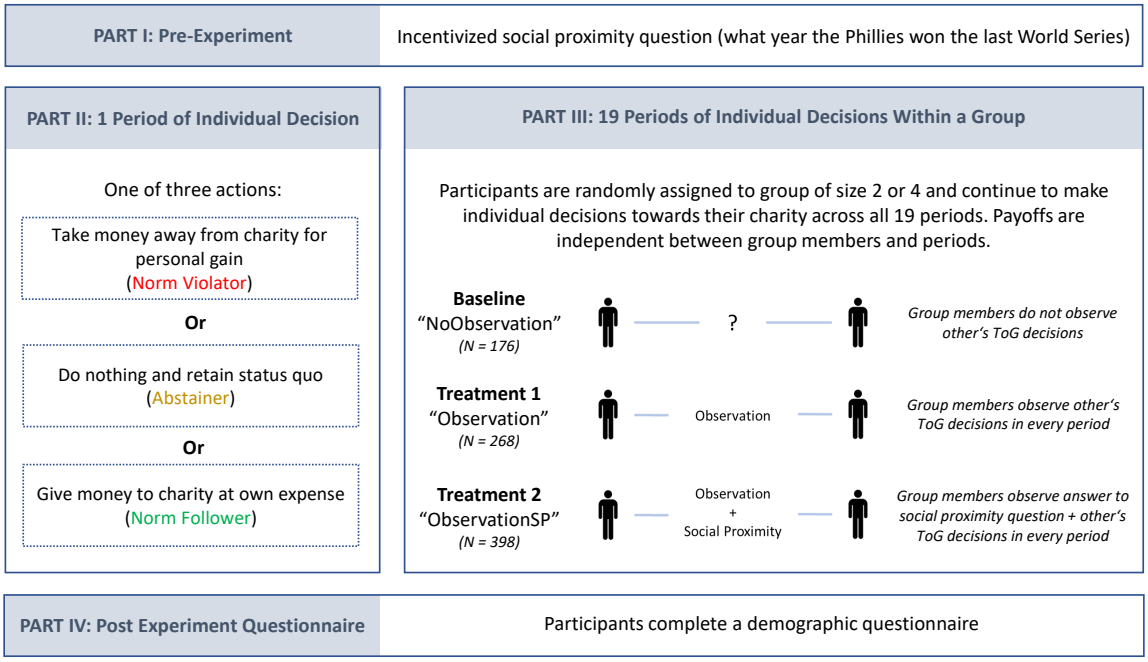


Figure 1: Experimental Design.

In each treatment, every participant went through the same four parts of the experiment (plus a payment phase). 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) and \$0 otherwise. Overall, about 27% of our participants knew the correct answer to the question. In the ObservationSP treatment (for details, see below), the answers to the knowledge question were used to prime social proximity: at the end of Part II, participants were matched in groups and informed about how each member of their group had answered the question (and thus about the extent to which they could be considered supporters of the Philadelphia team).[2] In the other treatments, participants were also asked the knowledge question, but were not informed about the answers of the other group members. In these treatments, participants learned whether their answer to the question was correct or not (and hence their payment for it) at the end of the experiment, while in the ObservationSP treatment, this information was given between Part II and Part III.

Part II - Individual Decision (one period)

Next, all participants made one ToG decision. As explained in Section 3.1, 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 another participant's payoff or the payoff of a charity besides the individual's own. This removes any type of interdependency and strategic considerations between participants' payoffs.

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 and was solely done for the purpose of whose decisions subjects could observe. We varied group size because existing evidence suggests that group size affects the pressure to conform

[2] In order 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 pro-social behavior in other, unrelated strategic contexts (Depetris-Chauvin et al., 2020). We elicit this type of social identification by asking subjects about their knowledge of an important sports event. Those that know about it (the victory of a team) are likely to be fans of that team. Part of our contribution is to assess if this type of identification also elicits positive behavior in a non-strategic context. Whether this also induces norm following is an empirical question we answer as well. An advantage of using a knowledge question rather than a preference question ("Are you a fan of this sports team?") is the mitigation of a social desirability bias and demand effects.

with group norms (see Bond, 2005). Each participant continued to make the same type of individual decision as in Part II towards the charity for another 19 periods. In each period, both the subject and the charity were endowed with 100 ECUs each in order to avoid any path-dependent carryover effects from previous periods; payoffs were also strictly individual and hence independent between group members. 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 empirical information has on compliance to the giving norm in the ToG donation 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 in 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 with respect to the giving norm is 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 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 I, 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, strategic substitution effects ('I can take because you give and balance out the payoffs'), among others. 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.

3.4. Behavioral Hypotheses

In this section, we present hypotheses derived from the conceptual framework outlined in Section 2. First, we conjecture that a norm of giving to charities exists in our experimental setting. A necessary condition for the existence of a norm in the ToG donation game is that people's normative expectations differ across the three options of taking, abstaining, and giving (if not, all three options would be believed to be normatively equally appealing, and hence there would be no norm). It is plausible to expect that fewer people will presume taking, rather than abstaining or giving, to be the most normatively-appealing option for the majority of participants (normative expectations). Another necessary condition is the existence of empirical expectations: people must expect a majority to abide by the norm.

H_1: *A norm of giving exists in the ToG donation game. Taking from the charity is considered the normatively least appealing and also least frequently expected option.*

Having established that a norm of giving exists in the ToG donation game and that a majority of people think that both giving is expected *and* normatively desirable, we then study how the availability of empirical information affects compliance with it. As discussed in Section 2, we expect an asymmetric effect of empirical information on norm compliance: evidence of norm-violating ToG decisions will have a stronger impact than evidence of norm compliance. Consequently, we expect a decline in compliance with the

giving norm across the 19 periods of Part III in Observation, compared to NoObservation where information is not available. Thus, our second hypothesis is:

H₂: *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.*

As discussed in Section 2, we conjecture that the asymmetric effect of empirical 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 also expect social proximity to reduce this self-serving use of empirical information: 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 socially proximate groups. This conjecture implies that we should expect a smaller decline in norm compliance in ObservationSP compared to Observation. Moreover, we expect that this effect is driven by the most socially proximate groups in ObservationSP, i.e. the groups where all members have given the same answer to the sport team's social proximity question of Part I of the experiment.

H₃: *Social proximity reduces the asymmetric effect of empirical 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 socially proximate groups.*

4. Results

4.1. Norm-Elicitation Experiment: Existence of a Norm of Giving to Charity

Figure 2 shows the results of the norm-elicitation experiment, where we elicited subjects' personal normative beliefs, normative expectations, and empirical expectations. 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 that imply taking from the charity ("Take"), actions that leave the charity's endowment untouched ("Abstain"), and actions that imply giving to the charity ("Give").

Subjects' personal normative beliefs are shown by the black solid line in Figure 2, representing the percentage of subjects ($N = 108$) who personally believe that one ought to

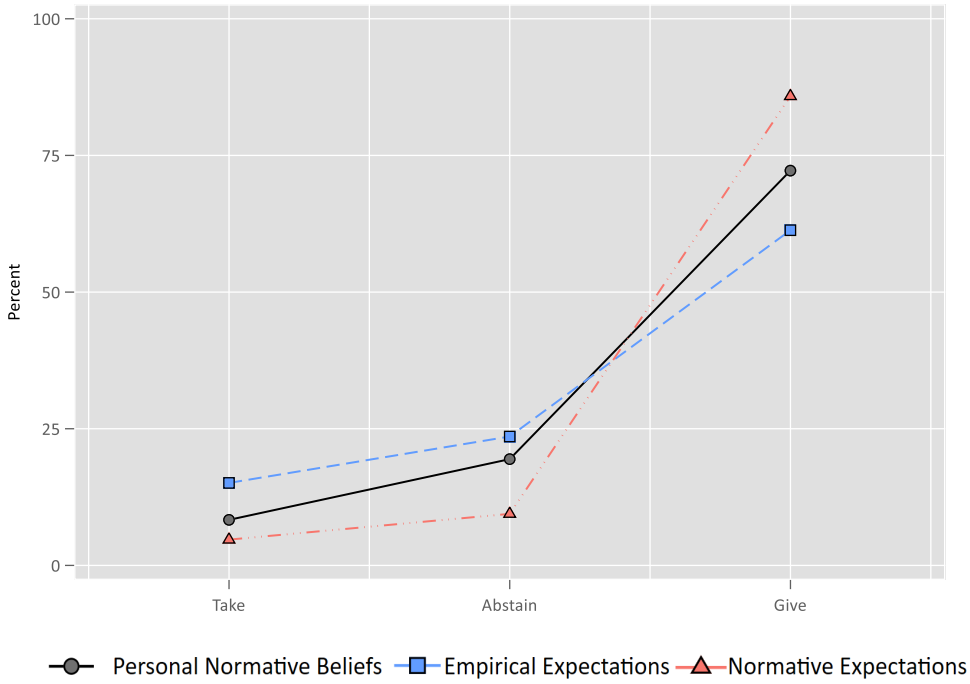


Figure 2: Personal normative beliefs, normative expectations and empirical expectations in the ToG game.

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 2 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 normative beliefs. Together, the personal normative 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 H_1 and shows that normative expectations that prescribe giving and proscribe taking exist in the ToG donation game. The fact that they are also consistent with personal normative 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 2 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 before 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.[3] 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.

Overall, the norm-elicitation experiment indicates that a norm of giving exists in the ToG game, confirming H_1. Showing that a norm exists, however, does not necessarily imply that it will be followed. In our experiment, material incentives support selfish behavior, because choices are anonymous and no sanctions are present. In the next section, we will explore how compliance with the norm of giving unfolds in the behavioral experiment.

4.2. Behavioral Experiment

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

We observe the donation behavior of $N = 842$ subjects across 20 periods of the ToG game. The first period of the game (Part II of the experiment) is exactly the same in all of our 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 is followed when subjects have to rely on their homegrown expectations about what ought to be done and what is commonly done.

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 (details are in the Online Appendix). The table also displays the percentage of participants

[3] 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.7 in the Online Appendix.

	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.

in each treatment who gave to, took from, 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".

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 personal, normative, and empirical beliefs 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 is not entirely surprising in light of the consideration that a norm may exist and not be followed at a given time. In fact, the actual level of taking and abstention in the behavioral experiment is much greater than what was anticipated by subjects in the norm-elicitation experiment. Taken together, these results point to the importance of differentiating between the existence of a norm and the norm being followed at any given time in a population.

4.2.2. 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 empirical expectations downwards and towards the actual behavior in the game. Figure 3 (left panel) shows

the evolution of average donations over the 19 periods of Part III across our treatments. Figure 3 (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 3). A positive value indicates that, relative to the average in Part II, the amount given to the charity has increased, and a negative value indicates that the amount taken from the charity has increased.

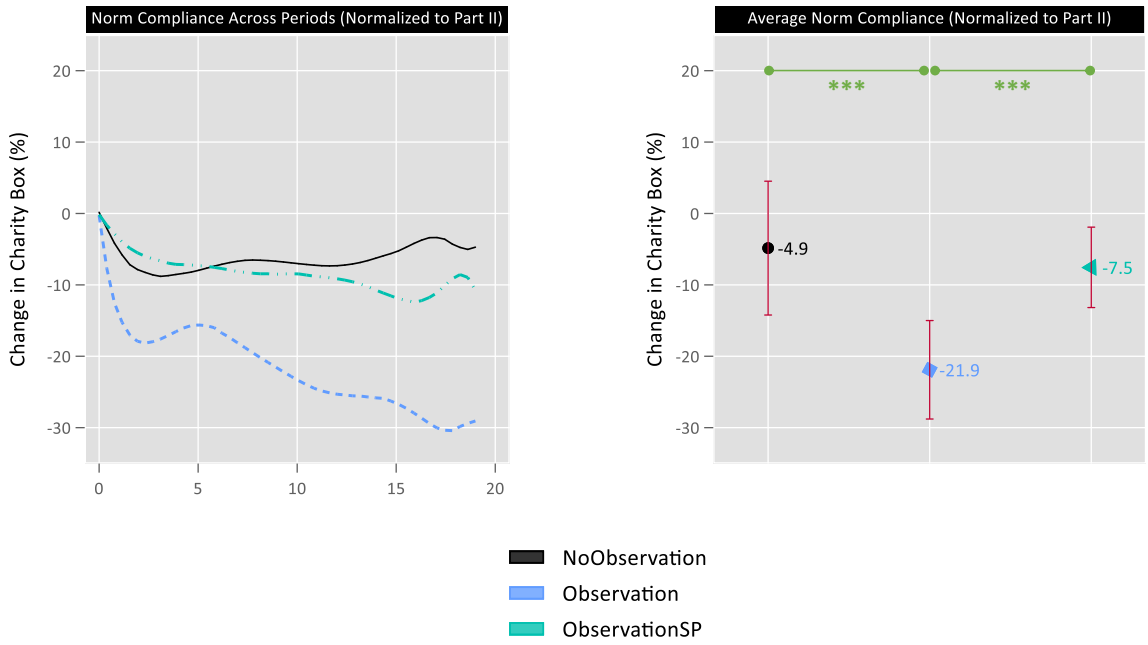


Figure 3: Left panel: Subject's behavior across treatments and periods in Part III, normalized to average behavior in Part II (period 0 in the graph). For purposes of readability, lines represent kernel-weighted local polynomial smoothing of degree 5. Right panel: Average subject's donations across treatments, 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. 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 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, participants took 4.9 additional ECUs from the charity 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 participants took on average only an additional 7.5 ECUs from the charity. 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 hypothesis 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 differences are detected between NoObservation and ObservationSP ($z = 0.877$, $p = 0.381$).[4]

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 NoObservation 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, participants in the Observation treatment took significantly more ECUs from the charity; the difference is significant at the 1% level. Moreover,

[4]Importantly, our main result is not driven by the disproportionate evidence of taking-behavior in Part II. In fact, as illustrated in Figure A.1 in the Online Appendix, we observe a substantial decay in norm compliance even for the sub-groups in which such behavior was initially in the minority. This additional robustness check emphasizes the generalizability of the observed dynamics.

a test of equality of coefficients between Observation and ObservationSP confirms that participants took more from the charity 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 second hypothesis, H_2, about the effects of empirical information. 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_3.

Why does norm compliance not significantly decline when information about social proximity is available? At the heart of H_2 and H_3 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 learning within groups.[5]

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

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

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.

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 empirical information in Observation relative to NoObservation.[6] 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.[7] In contrast, observing norm-violating behavior has a strong and negative effect on donations. We estimate that receiving feedback that other group members take 1 ECU from the charity reduces the amount donated on average by -0.29 + 0.05 = -0.24 ECUs (see Table 2, column 2). This effect is significant at the 1% level.[8] An F-test shows that the effects of examples of violation are statistically significantly larger than the effects of examples of compliance ($\tilde{\chi}^2(1) = 18.41$, $p < 0.001$).

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 of 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.[9] 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). This

[6] 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 social 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' behaviors did not have any significant effect on donations.

[7] We cannot reject that *Pos. Avg Don. t-1 + Pos. Avg Don. t-1 ∗ Observation = 0* ($p = 0.87$) or that *Zero Avg Don. t-1 + Zero Avg Don. t-1 ∗ Observation = 0* ($p = 0.81$).

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

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

effect is significant at the 1% level.[10] 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.[11] 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).

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 induce 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 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_3, 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

[10] We reject: *Neg. Avg Don. t-1 + Neg. Avg Don. t-1 * ObservationSP = 0* ($\tilde{\chi}^2(1) = 13.23$, p < 0.001).
[11] We reject: *Pos. Avg Don. t-1 + Pos. Avg Don. t-1 * ObservationSP = 0* ($\tilde{\chi}^2(1) = 8.02$, p = 0.005) and both interactions (-0.02 and 0.21) are significantly different from each other ($\tilde{\chi}^2(1) = 30.70$, p<0.001).

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.

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.[12]

4.2.3. Heterogeneity in the Erosion of Norm Compliance

An additional question that we address in this subsection is whether initial dispositions toward norm compliance mediate the responsiveness 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 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 they continuously received empirical information about others' ToG decisions in their group. Figure 4 plots separately for each type of subject and treatment, with 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 a different breakdown of the data see Figure A.5 in the Online Appendix).

We have a number of interesting results. In NoObservation (leftmost column of Figure 4), most groups only make small changes to their behavior between Part II and III of the experiment, which is to be expected given that participants neither observe nor are being observed by others. Norm followers give on average 65.3 ECUs to charity in Part II and 52.2 ECUs in Part III, a decrease of 13.1 ECUs that is significant at the 5 percent level (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

[12] 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 Appendix Figures A.2 and A.3. The suggestive evidence is in line with our theoretical discussion.

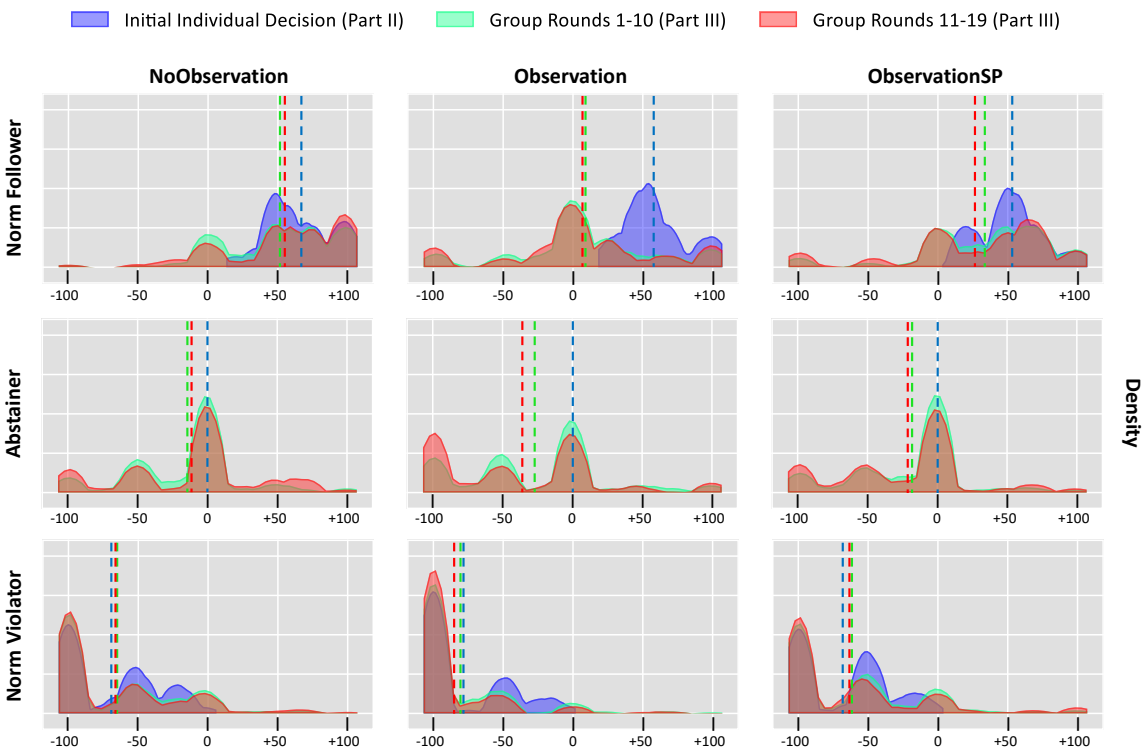


Figure 4: Shown are the distribution of donations (density) disaggregated by group of subject (norm violators, abstainers, norm followers, as defined by Part II initial behavior) and treatment. Vertical lines represent averages. Colors correspond to the respective phases. Blue: donations in Part II. Green: average donations across periods 1-10 of Part III. Red: average donations across periods 11-19 of Part III.

Part III is 64.9 ECUs, which is not statistically significant (MWU, $p = 0.140$). The change in behavior of norm followers and abstainers is larger than that of norm violators (both $p < 0.005$), while we find no difference between norm followers and abstainers ($p = 0.994$).[13]

In Observation (middle column of Figure 4), 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

[13]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).

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 4), 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 find no difference 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 abstainers, and then initial norm violators. The reason for this is straightforward: based on Figure 3, 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 framework presented in Section 2: we expect norm followers to be those who would have to make the largest (downwards) revisions to their empirical expectations after receiving information about how other members of their group actually behaved. Since they have a conditional preference for following the norm, their behavior is strongly influenced by what they observe.[14]

4.2.4. The Mechanisms Behind the Erosion of Norm Compliance

In this final subsection, we conclude with a discussion of the possible deeper mechanisms that may underlie the observed erosion of norm compliance in the Observation treatments of the behavioral experiment. As we discuss at length in Section 2, a first *direct* mechanism lies in the updating of empirical expectations: participants' initial expectations about others' behavior are inevitably affected by observation of what others actually do

[14]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). See Figure A.6 in the Online Appendix.

in the experiment. Since, in our theoretical framework, norm compliance is conditional on empirical expectations, a downward revision of empirical expectations—triggered by observation of norm violations—implies a drop in norm compliance. Analogously, observation of norm compliance should lead to an upward revision of empirical expectations and induce further compliance. The fact that examples of violations and compliance have asymmetric effects in Observation suggests that the process of belief updating may be biased towards self-interest, in the absence of additional motivations to pay attention to signals of compliance. As discussed in Section 2 and shown above, social proximity provides the necessary additional incentives to respond also to examples of norm compliance.

Observing others' behavior may also affect compliance *indirectly*, 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 revise their normative expectations about what others think one ought to do in the game. The indirect effect of observation on normative expectations has been documented in other settings, e.g., Bicchieri et al. (2019a) and Gächter et al. (2019). Whether this type of revision also affects personal normative beliefs is an empirical question. Though we do not believe this to be the case, we also test for this possible effect.

To test whether an indirect effect is obtained in our setting, we conducted an additional follow-up experiment, where we elicited participants' personal normative beliefs and normative expectations in the ToG game *twice*: before and after being shown how participants in the behavioral experiment actually behaved. The experiment was conducted with $N = 174$ students at the University of Pennsylvania who had not participated in any of our previous experiments (we test the robustness of our results on Amazon Mechanical Turk and find qualitatively similar results, see Online Appendix for details). Participants received \$2 upfront for participating and up to \$4 if they answered the belief questions correctly (see discussion below). The average duration of the experiment was 10 minutes. The procedures of this "norm-updating" experiment were similar to those of the norm-elicitation experiment described in section 3.2 (see Online Appendix for more details). To elicit personal normative 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 3 (left

panel). Participants were randomly shown behavior in only one of the three treatments (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 normative beliefs and their (incentivized) normative expectations a second time.

We present our results in Figure 5 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).

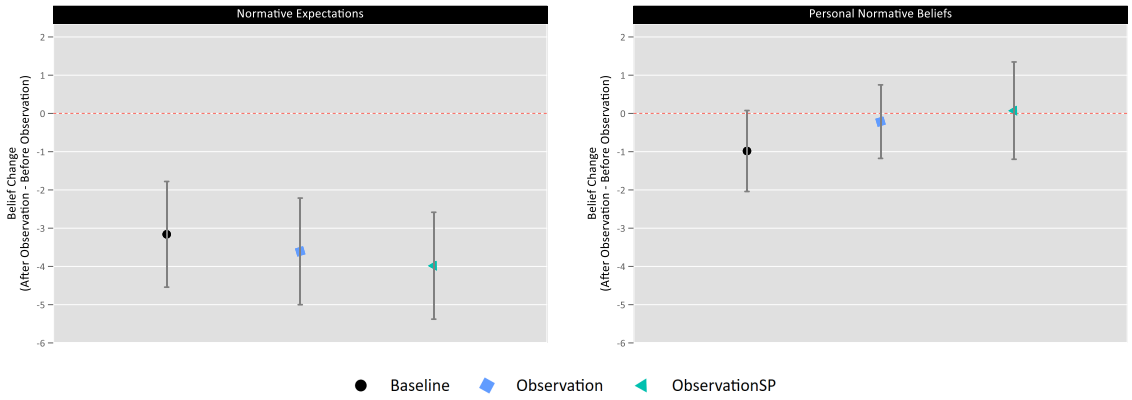


Figure 5: Left panel: Changes in normative expectations. Right panel: Changes in personal normative beliefs across treatments. Negative (positive) values indicate that the beliefs after the observation of behavior were lower (higher) than beliefs prior to observation of behavior. Whiskers represent 95% CIs.

We find evidence of a significant indirect effect of observing behavior on normative expectations. In all cases, normative expectations (left panel Figure 5) were adjusted significantly downwards after observing evidence of violations of the norm of giving (all p-values < 0.01). For personal normative beliefs (right panel Figure 5), we observe small and insignificant revisions of beliefs. The differential adjustment of personal normative beliefs and normative expectations in light of updated empirical expectations is consistent with our theory discussed above: preferences are conditional on empirical and normative expectations, whereas personal normative beliefs are rather unconditional value assessments.

We would therefore expect greater stability in personal normative beliefs than normative expectations. Overall, these results suggest that observing others' behavior has both a direct and an indirect effect on norm compliance, as observation leads subjects to update both their empirical and normative expectations. Since norm compliance is conditional on such expectations, revisions in these beliefs trigger a change in behavior. Our results show that the process of norm-updating and behavioral change can quickly lead to the erosion of compliance in the absence of additional factors that motivate individuals to respond to examples of both norm compliance and norm violations, such as social proximity.

5. 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 norm compliance. To investigate norm compliance, we adopted the theory of social norms developed by Bicchieri (2006, 2016) and designed a novel, non-strategic Take-or-Give donation game, where people could give to charity, take from it, or abstain from changing the initial allocation between self and the charity. We first established that most people think taking from the charity is socially inappropriate, whereas giving to the charity is the social norm. Our behavioral results reveal a strong asymmetry in norm compliance: observing an anonymous other violating the norm (taking from the charity) increased the likelihood that people 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 others eroded norm compliance in a way that was not compensated by observing norm compliance. No erosion occurred when others' behavior was unobservable.

While this asymmetry in reactions paints a bleak picture about norm compliance when anonymous others can be observed, in most real-world interactions people have clues about their social proximity to the people they observe. Our experiment accounted for this fact by designing a treatment where a minimal cue about social proximity, fanhood of a sports team, 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 norm compliant behavior, thus halting norm erosion. Our results on 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 when judging one's own actions compared to an out-group's (Gino et al., 2009b). Our novel contribution is to show that social proximity can stabilize norm compliance by reducing the asymmetry in reactions to observing norm violations and norm compliance. Punishment, often seen as an important stabilizer of social norms, may not be necessary when social proximity induces emulation of norm compliance.

Our results have several key implications for understanding the dynamics of norm compliance and for designing behavioral change interventions (for a discussion, see Bicchieri and Dimant, 2019). First, in order to predict compliance, it is essential to measure not only what people regard as appropriate behavior in a given situation (e.g., by using the norm-elicitation technique by Krupka and Weber, 2013), but also to gauge the extent to which people believe others will actually engage in that behavior. In this paper, building on Bicchieri and Chavez (2010), we propose a set of belief-elicitation measures to accomplish this. Second, we show that an indirect effect of observing norm transgressions is a change in normative expectations. The result is that the transgression appears more acceptable; if not to oneself, at least to others. In the long run, a lasting change in empirical *and* normative expectations would result in norm erosion. Finally, our results point to 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, our study shows that this comes at the cost of missing important insights into the drivers of norm compliance. For example, 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. Most social encounters in natural and online environments involve observation of shared characteristics, thereby inducing social categorization and group identification. Information about others' behavior has a more balanced effect in natural settings where signals of social proximity are available, compared to our Observation treatment that deliberately suppresses any signal of social proximity. Effective policy interventions should take into account that people are likely to pay attention to norm compliance by similar others, less so by anonymous or otherwise generic others.

References

Abbink, K., Gangadharan, L., Handfield, T., and Thrasher, J. (2017). Peer punishment promotes enforcement of bad social norms. *Nature Communications*, 8(1):609.

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

Allen, V. L. (1965). Conformity and the role of deviant. *Journal of Personality*, 33(4):584–597.

Andreoni, J., Nikiforakis, N., and Siegenthaler, S. (2017). Social change and the conformity trap. Technical report, Mimeo.

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.

Bartling, B. and Özdemir, Y. (2017). The limits to moral erosion in markets: Social norms and the replacement excuse. CESifo Working Paper Series.

Bartling, B. and Schmidt, K. M. (2015). Reference points, social norms, and fairness in contract renegotiations. *Journal of the European Economic Association*, 13(1):98–129.

Bašić, Z., Falk, A., and Kosse, F. (2019). The development of egalitarian norm enforcement in childhood and adolescence. *Journal of Economic Behavior & Organization*.

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. (2019a). 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. (2019b). 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.

Bicchieri, C. and Xiao, E. (2009). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, 22(2):191–208.

Bohnet, I. and Zeckhauser, R. (2004). Social comparisons in ultimatum bargaining. *Scandinavian Journal of Economics*, 106(3):495–510.

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.

Bott, K. M., Cappelen, A. W., Sorensen, E., and Tungodden, B. (2019). You've got mail: A randomised field experiment on tax evasion. *Management Science*.

Brennan, G., Pettit, P., et al. (2004). *The economy of esteem: An essay on civil and political society*. OUP Oxford.

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.

Burke, M. A. and Young, H. P. (2011). Social norms. In *Handbook of Social Economics*, volume 1, pages 311–338. Elsevier.

Bursztyn, L., Egorov, G., and Fiorin, S. (2019). From extreme to mainstream: How social norms unravel. NBER Working Paper.

Centola, D. (2018). *How behavior spreads: The science of complex contagions*, volume 3. Princeton University Press.

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.

Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review*, 108(10):3028–56.

Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika*, 72(2):263.

Cialdini, R. B. and Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55:591–621.

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

Croson, R. and Shang, J. Y. (2008). The impact of downward social information on contribution decisions. *Experimental Economics*, 11(3):221–233.

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.

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

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., 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.

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.

Elster, J. (1989). *The cement of society: A survey of social order*. Cambridge University Press.

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

Falk, A., Neuber, T., and Szech, N. (2020). Diffusion of Being Pivotal and Immoral Outcomes. *The Review of Economic Studies*. rdz064.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives*, 14(3):159–181.

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

Fellner, G., Sausgruber, R., and Traxler, C. (2013). Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association*, 11(3):634–660.

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. (2019). The behavioral logic of rule following and social norm compliance. Unpublished Manuscript.

Gächter, S., Nosenzo, D., and Sefton, M. (2013). Peer effects in pro-social behavior: Social norms or social preferences? *Journal of the European Economic Association*, 11(3):548–573.

Gerber, A. S., Green, D. P., and Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, 102(1):33–48.

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.

Goldin, C. and Katz, L. F. (2002). The power of the pill: Oral contraceptives and women's career and marriage decisions. *Journal of Political Economy*, 110(4):730–770.

Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3):472–482.

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

Hallsworth, M., List, J. A., Metcalfe, R. D., and Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, 148:14–31.

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

Jachimowicz, J. M., Hauser, O. P., O'Brien, J. D., Sherman, E., and Galinsky, A. D. (2018). The critical role of second-order normative beliefs in predicting energy conservation. *Nature Human Behaviour*, 2(10):757.

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.

Kling, J. R., Ludwig, J., and Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics*, 120(1):87–130.

Knack, S. (1992). Civic norms, social sanctions, and voter turnout. *Rationality and Society*, 4(2):133–156.

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. 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.

Lindbeck, A., Nyberg, S., and Weibull, J. W. (1999). Social norms and economic incentives in the welfare state. *The Quarterly Journal of Economics*, 114(1):1–35.

Lindström, B., Jangard, S., Selbing, I., and Olsson, A. (2018). The role of a "common is moral" heuristic in the stability and change of moral norms. *Journal of Experimental Psychology: General*, 147(2):228.

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

McDonald, I. M., Nikiforakis, N., Olekalns, N., and Sibly, H. (2013). Social comparisons and reference group formation: Some experimental evidence. *Games and Economic Behavior*, 79:75–89.

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.

Muthukrishna, M., Francois, P., Pourahmadi, S., and Henrich, J. (2017). Corrupting cooperation and how anti-corruption strategies may backfire. *Nature Human Behaviour*, 1(7):0138.

Neugebauer, T., Perote, J., Schmidt, U., and Loos, M. (2009). Selfish-biased conditional cooperation: On the decline of contributions in repeated public goods experiments. *Journal of Economic Psychology*, 30(1):52–60.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14(3):137–158.

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

Pryor, C., Perfors, A., and Howe, P. D. (2019). Even arbitrary norms influence moral decision-making. *Nature Human Behaviour*, 3(1):57–62.

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.

Smerdon, D., Offerman, T., and Gneezy, U. (2019). Everybody's doing it: On the emergence and persistence of bad social norms. *Experimental Economics*.

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.

Sugden, R. (1998). Normative expectations: the simultaneous evolution of institutions and norms. In Ben-Ner, A. and Putterman, L., editors, *Economics, Values, and Organization*, page 73–100. Cambridge University Press, Cambridge.

Sugden, R. (2000). Team preferences. *Economics & Philosophy*, 16(2):175–204.

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.

Turner, J. C. (1985). Social categorization and the self concept: A social cognitive theory of group behavior. In Lawler, E. J., editor, *Advances in Group Process*, volume 2, pages 77–122. JAI, Greenwich, CT.

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

Cristina Bicchieri, Eugen Dimant, Simon Gächter, Daniele Nosenzo

Contents

I. Robustness Checks

I.a. 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).

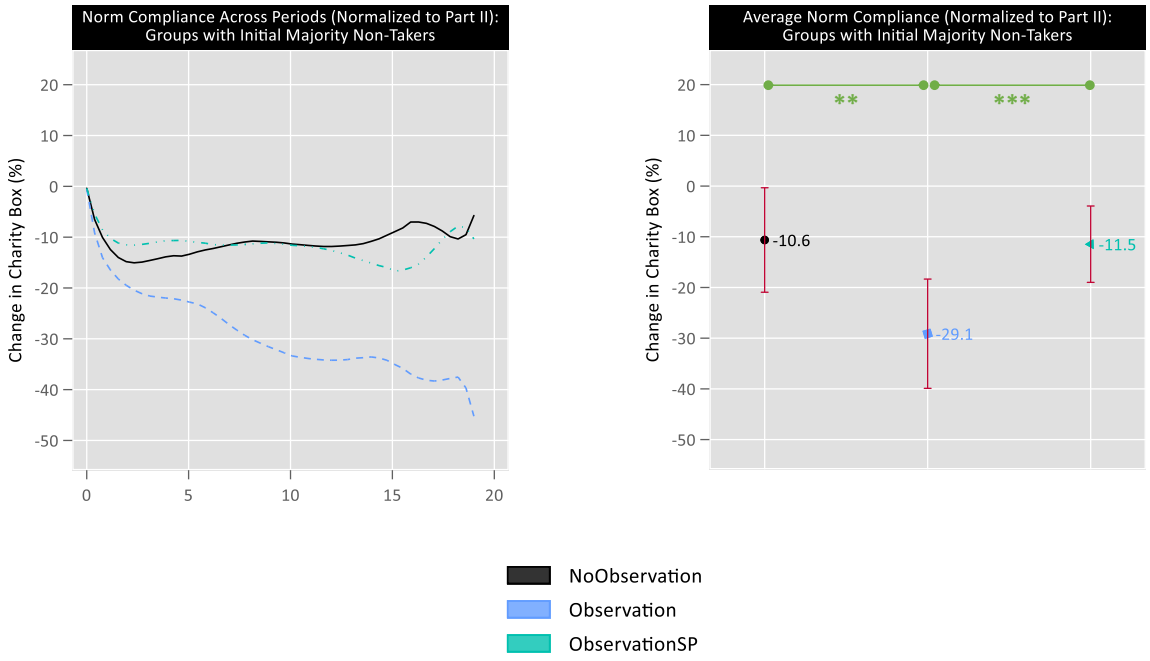


Figure A.1: Left panel: Subject's behavior across treatments and periods in Part III for groups in which the initial average behavior was not taking, normalized to average behavior in Part II (period 0 in the graph). For purposes of readability, lines represent kernel-weighted local polynomial smoothing 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.b. Group Proximity

Here we present graphical illustrations of the evolution of norm compliance in ObservationSP. Figure A.2 presents this for proximate groups, where all group members answered the proximity question in the same way (either all correctly or all incorrectly), as well as for non-proximate groups (mixed answers within the same group). Figure A.3 breaks down the data for the proximate 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.

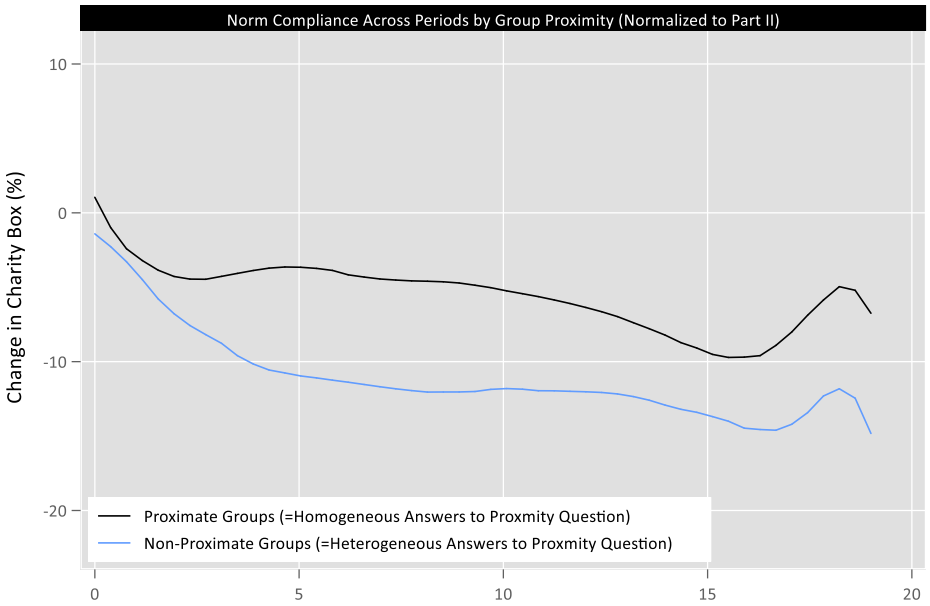


Figure A.2: Evolution of norm compliance (normalized to Part II behavior) in ObservationSP for groups answering the proximity question in the same way (proximate) or in different ways (non-proximate). Polynomial smoothing applied.

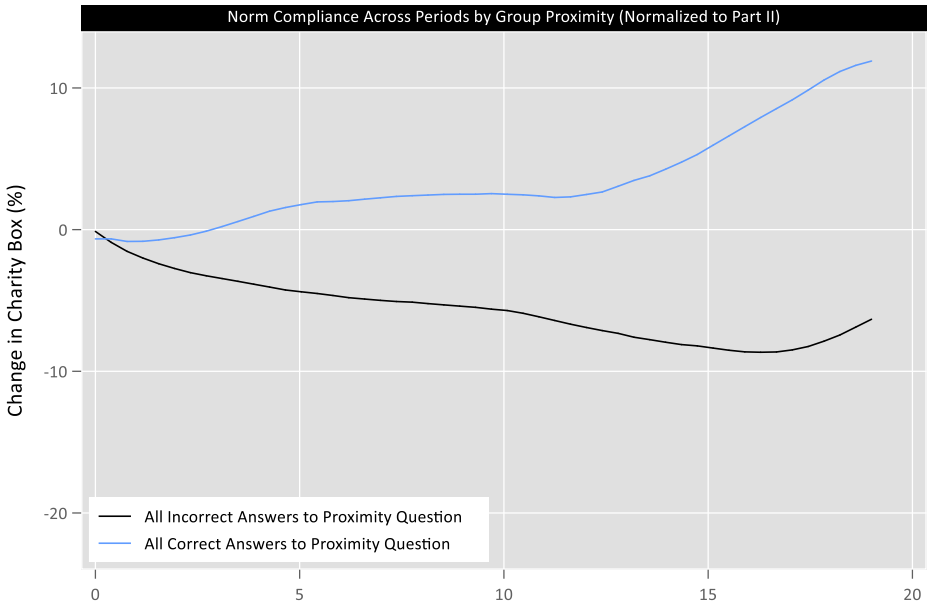


Figure A.3: Evolution of norm compliance (normalized to Part II behavior) in ObservationSP for groups answering the proximity question correctly or incorrectly. Polynomial smoothing applied.

I.c. Additional Heterogeneity Analysis

One additional way in which we can analyze the data from Section 4.2.3 is by looking at the stability of behavior across treatments. For this purpose, we use the same classification of initial behavior (initial norm violators, followers, abstainers) and examine the rate at which individuals switched from one category to the other across the different treatments. That is, conditional on the classification based on one's initial behavior, the following analysis sheds light on whether the frequency at which individuals switch to a different classification (e.g., from norm follower to norm violator) depends on the availability of information about others' behavior and proximity. The calculation is straightforward: for each period in the group stage, a participant's behavior is coded as '1' ('0') if the type of behavior (violator/follower/abstainer) was different (the same) as that participant's behavior in the initial donation decision in Part II. For each individual, we then average this sequence of 1s and 0s across all 19 periods, resulting in a single value per individual.

We present the results in Figure A.4. Our results show that individuals are substantially more likely to change behavior in Observation, which is highly statistically significantly different compared to NoObservation (test of proportions, $p < 0.01$) and weakly significantly different from ObservationSP (test of proportions, $p = 0.089$). The rate at which behavior changed in NoObservation and ObservationSP is not significantly different (test of proportions, $p = 0.12$). As indicated in the bottom panel, this is mainly driven by the initial norm followers and norm abstainers, whereas the likelihood that individuals who were initially violators change their type of behavior does not vary

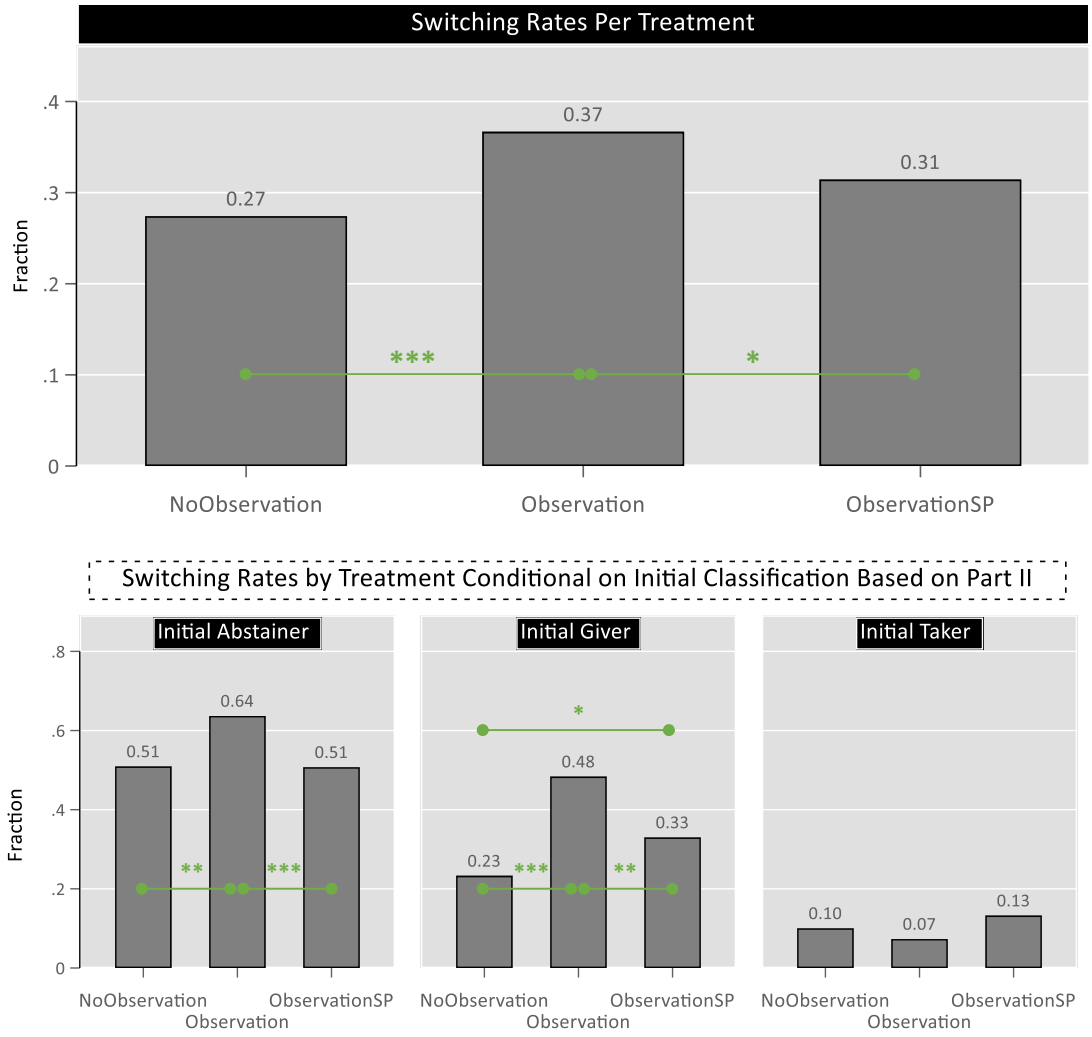


Figure A.4: Fraction of individuals switching their initial behavioral classification pooled across treatments (top panel) and conditional on the individual's initial behavior based on Part II (bottom panel). Stars indicate significant differences at the conventional levels of *p<0.1, **p<0.05, and ***p<0.01.

significantly across treatments (test of proportions, all $p > 0.16$). In more detail, initial abstainers are substantially more likely to switch to a different behavior classification in Observation compared to both NoObservation ($p = 0.02$) and ObservationSP ($p < 0.01$), whereas again the behavior in the latter two conditions is not statistically significantly different from each other ($p = 0.83$). Similarly, initial followers are substantially more likely to switch behavior in Observation condition compared to both NoObservation ($p < 0.01$) and ObservationSP ($p = 0.011$). Here, behavioral change in ObservationSP is marginally more likely than in NoObservation ($p = 0.098$).

Overall, these results suggest that with information about others' behavior, but not others' proximity, individuals are strongly influenced by feedback on how others behave and are substantially more likely to change their initial behavior. On the other hand, adding social proximity stabilizes behavioral change at levels similar to an environment devoid of information about others' behavior. At the same time, initial norm violators seem to be the most resistant to switching behavior and remain consistent throughout the game.

We can substantiate our results from Figure 4 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.5. 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.6). Note: the results do not change whether the cut-off values of -50 and +50 are included in the lower or upper half.

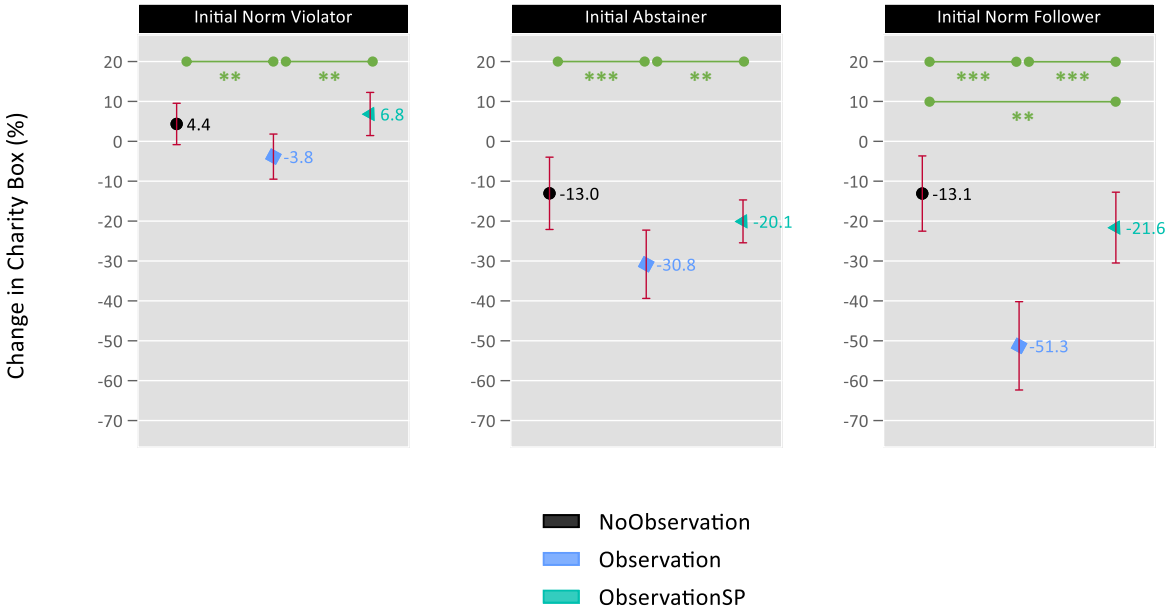


Figure A.5: Average subject's donations across treatments, averaged over the 19 periods of Part III conditional on their initial behavior in Part II. Whiskers indicate 95% confidence intervals. Stars indicate significant differences at the conventional levels of *p<0.1, **p<0.05, and ***p<0.01.

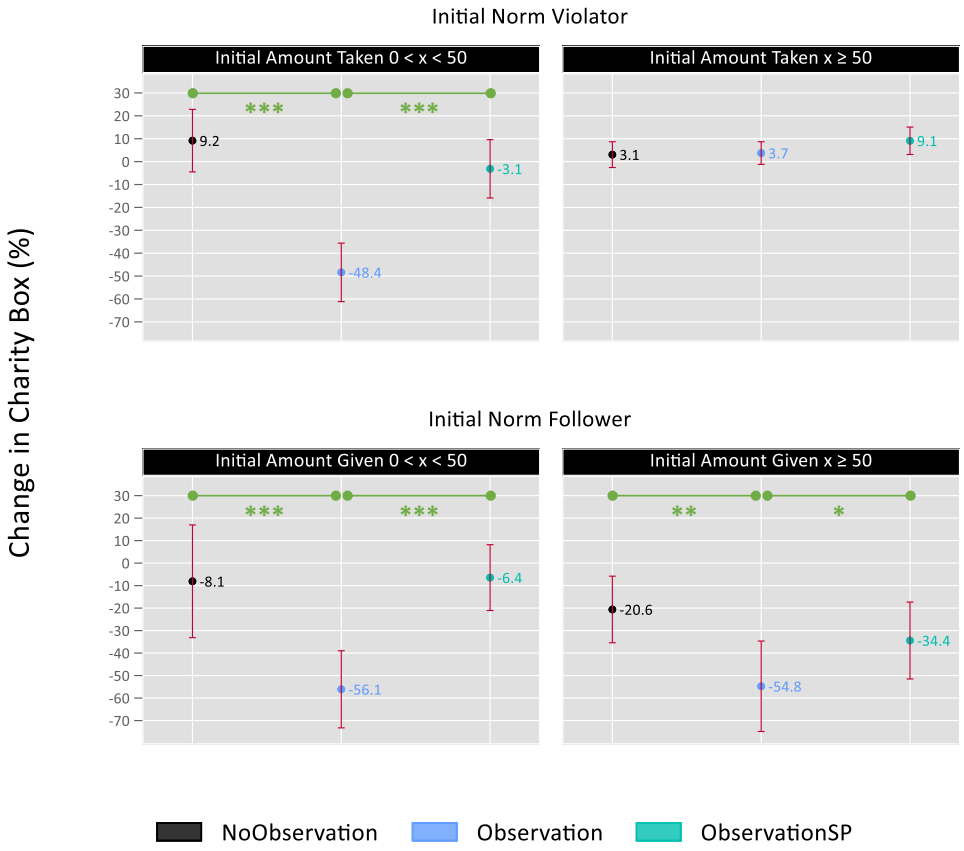


Figure A.6: Average subject's donations across treatments, averaged over the 19 periods of Part III conditional on their initial behavior in Part II and how extreme that behavior was (cut-off at +50 and -50, respectively). Whiskers indicate 95% confidence intervals. Stars indicate significant differences at the conventional levels of *p<0.1, **p<0.05, and ***p<0.01.

II. Laboratory Experiment

II.a. Instructions

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.
- 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.

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

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.)

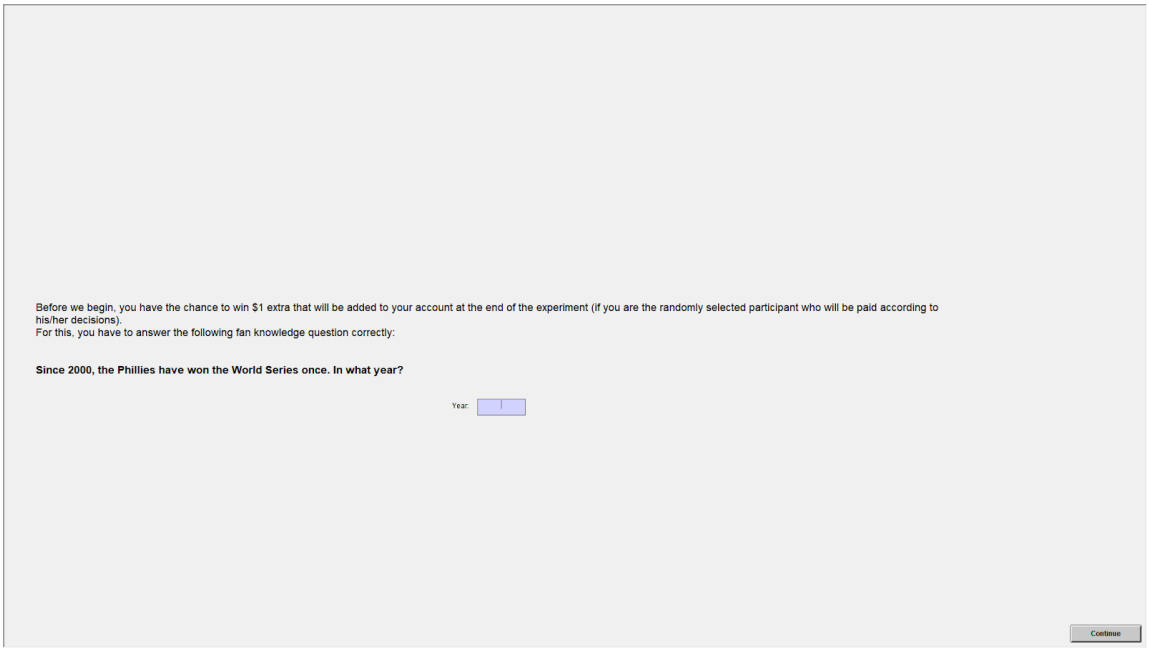
Payoff

- 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.
  - 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.

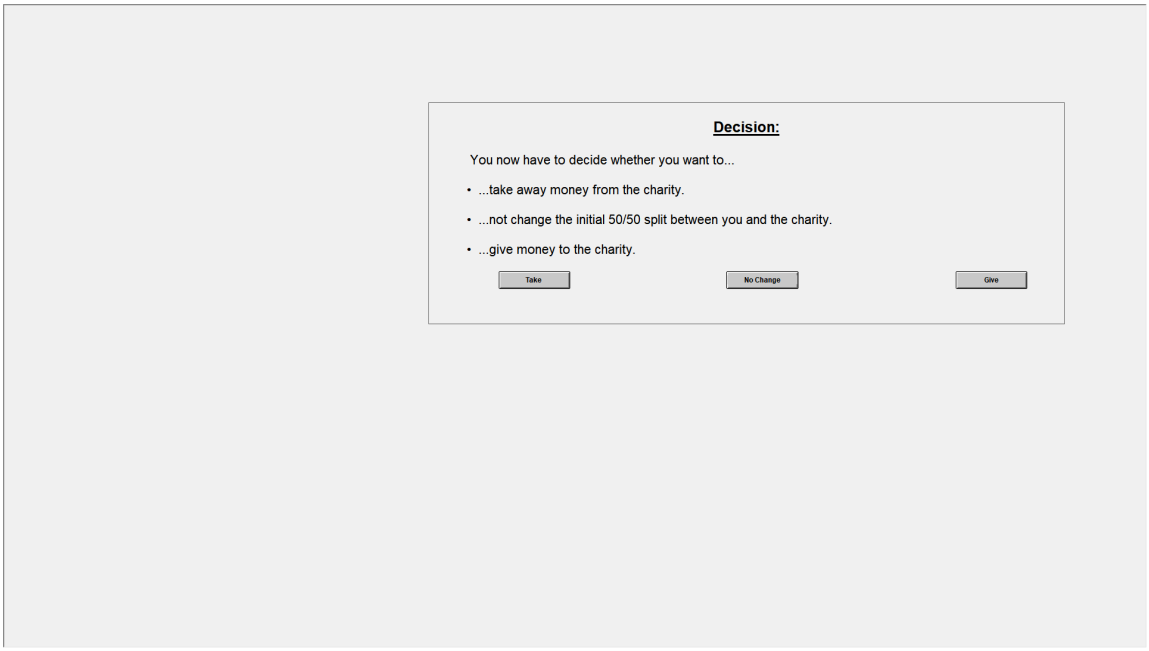
- The decisions that are made over the course of the next 19 rounds will remain anonymous. This means that individual behavior will not be displayed to any participant within your group or anyone else. (At the end of each round, every decision maker's decision will be revealed to every other decision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

II.b. Experimental Screenshots

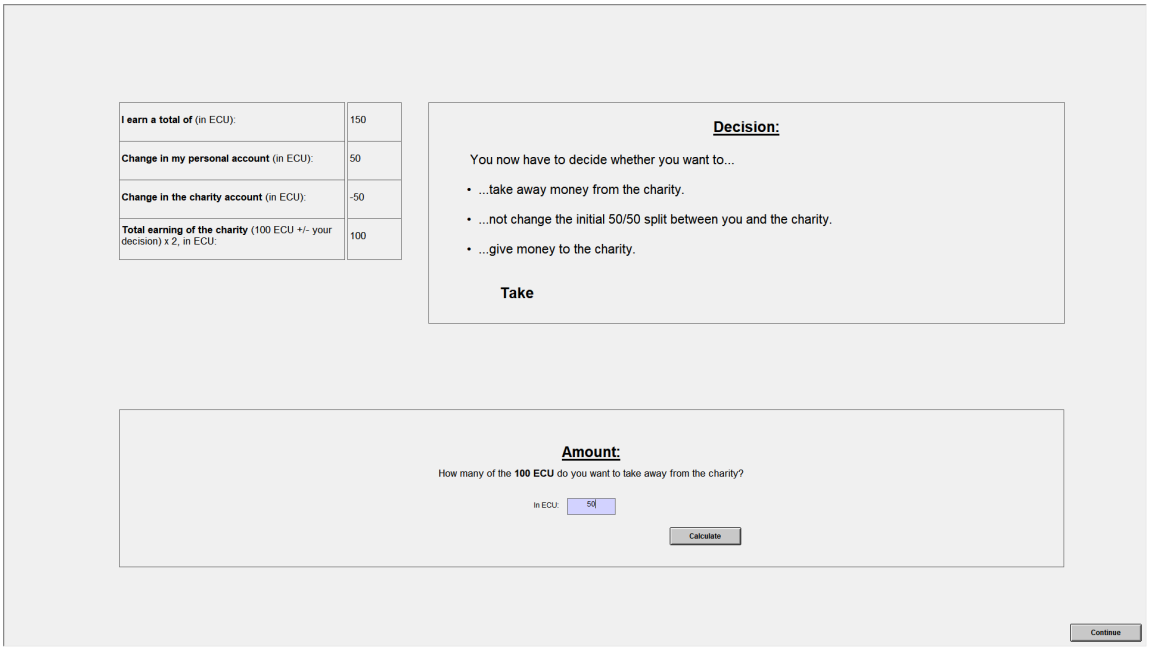
Here, we exemplarily present the screenshots for ObservationSP with group size 4. Screenshots for other treatments are available upon request. Merely Figure A5 changed across the different conditions: in NoObservation, participants did not see any information in other group members' history table. In 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.



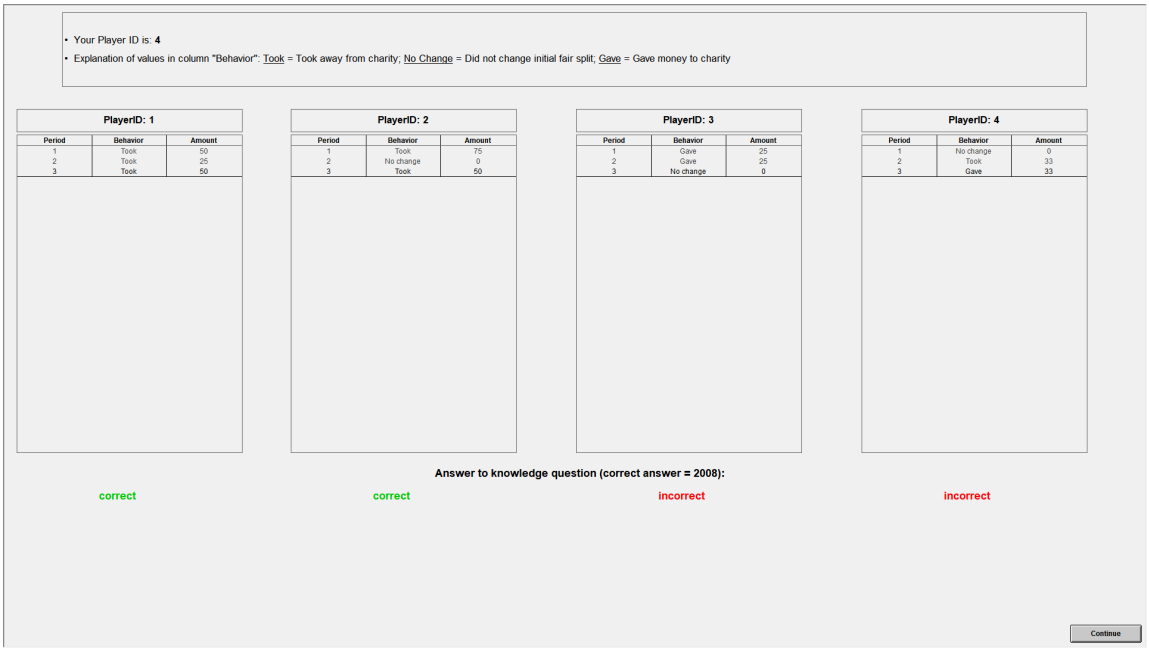
Screen 1: Knowledge question (correct answer: 2008).



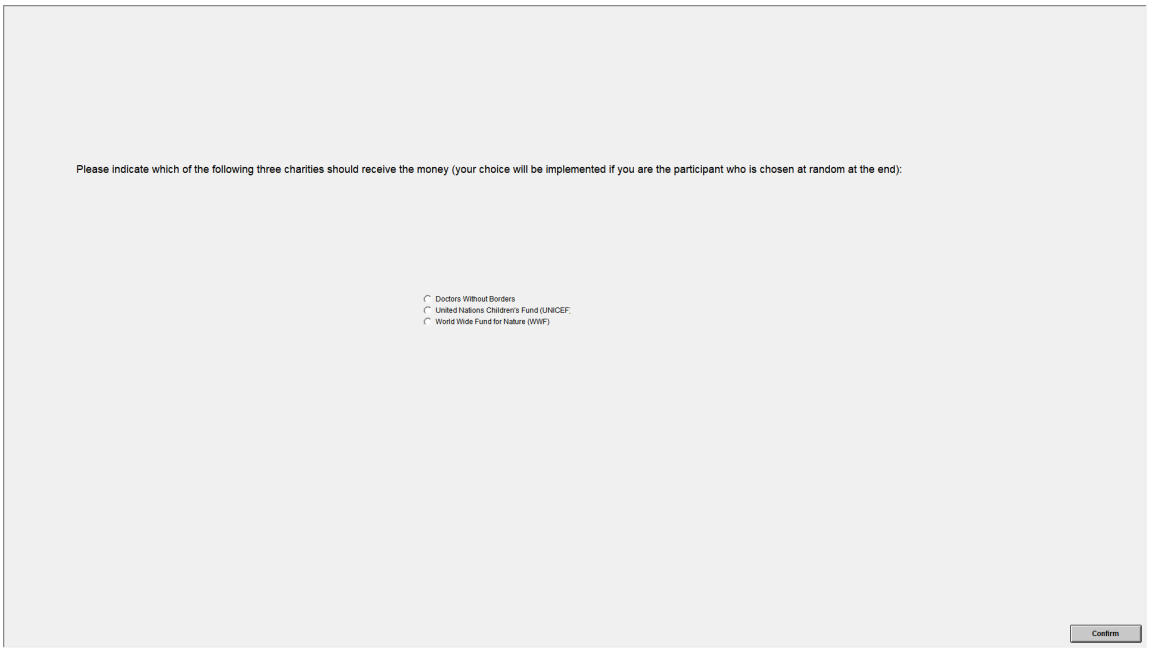
Screen 2a: Decision towards charity (choosing the type of behavior).



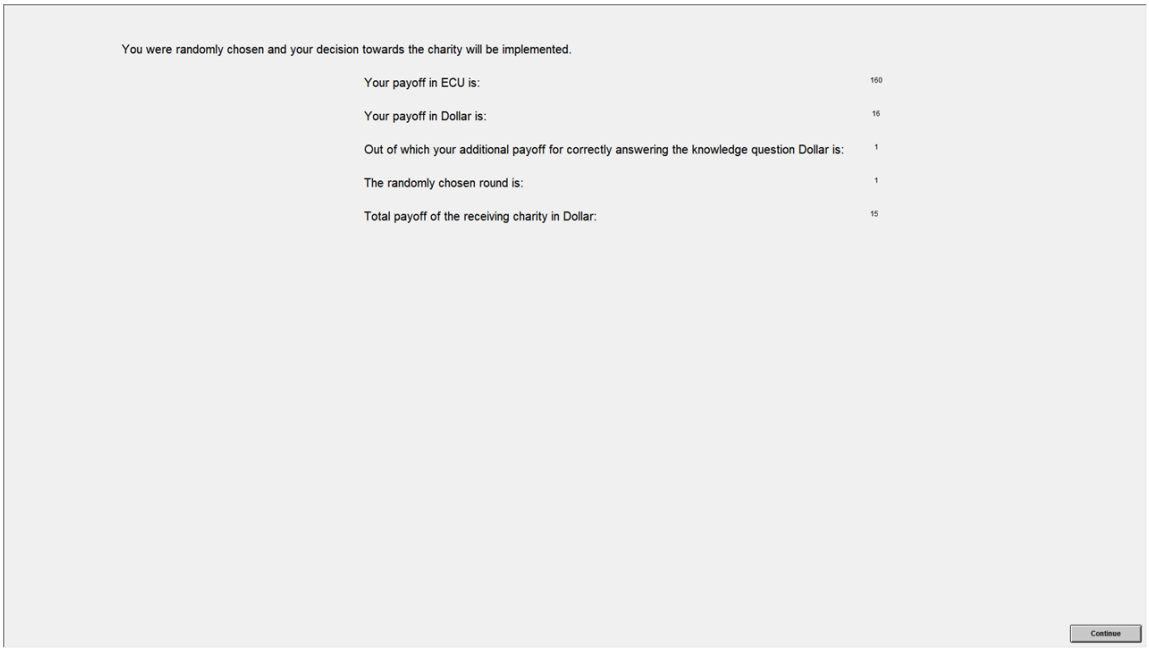
Screen 2b: Decision towards charity (choosing the amount).



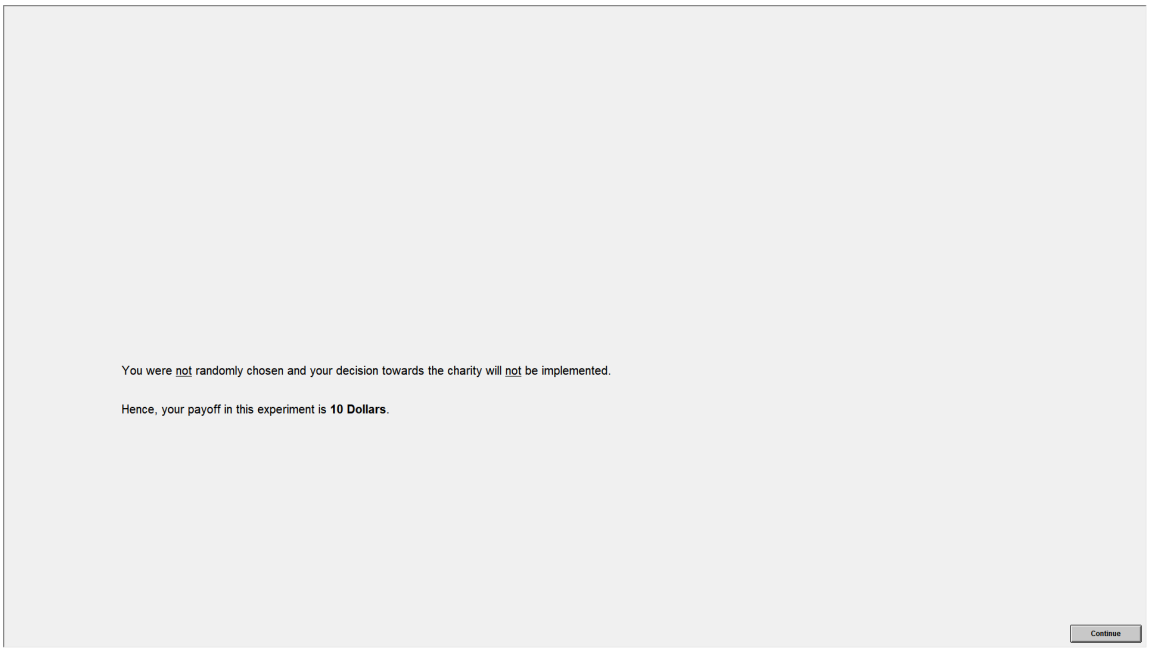
Screen 3: Observation stage (after 3 periods).



Screen 4: Observation stage (after 3 periods).



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



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

III. Norm Elicitation Experiments

III.a.1 Results Norm Elicitation Following Bicchieri and Chavez (2010)

Below, we present additional results and the screenshots from the norm elicitation experiments (all between-design) as discussed in the main text of the paper.

The figure below illustrates the distribution of beliefs across the various elicitation components. Details about the exact experimental mechanism can be found in Section 3.2 and additional results in Figure 2 the main text.

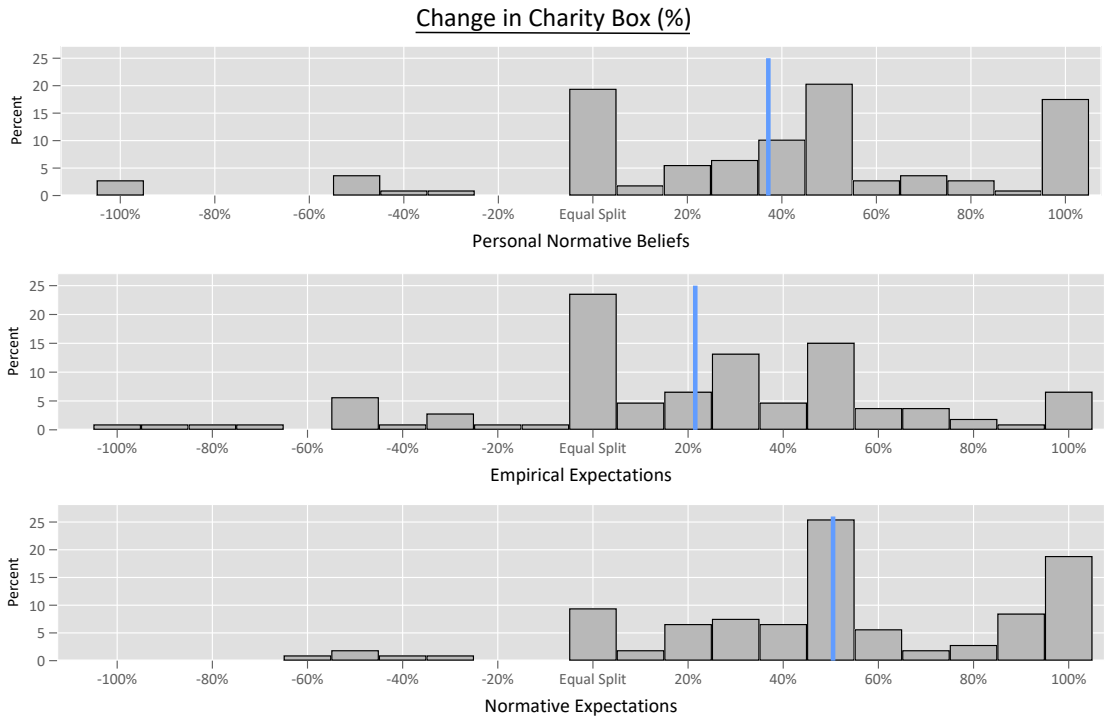
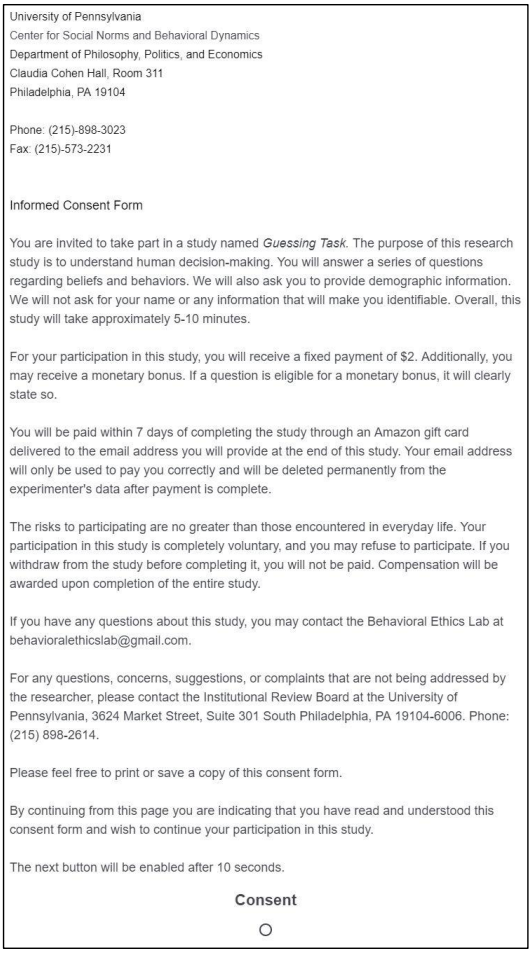


Figure A.7: Distribution of elicited beliefs across conditions. Vertical blue lines indicate averages.

III.a.2 Screenshots Norm Elicitation Experiment (Bicchieri and Chavez, 2010)

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.

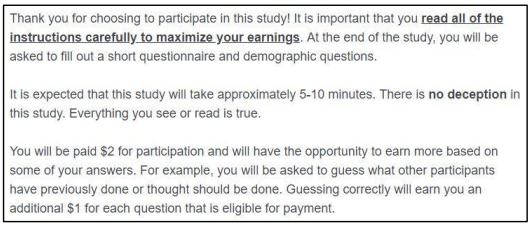
By continuing from this page you are indicating that you have read and understood this consent form and wish to continue your participation in this study.

The next button will be enabled after 10 seconds.

Consent

○

General Instructions

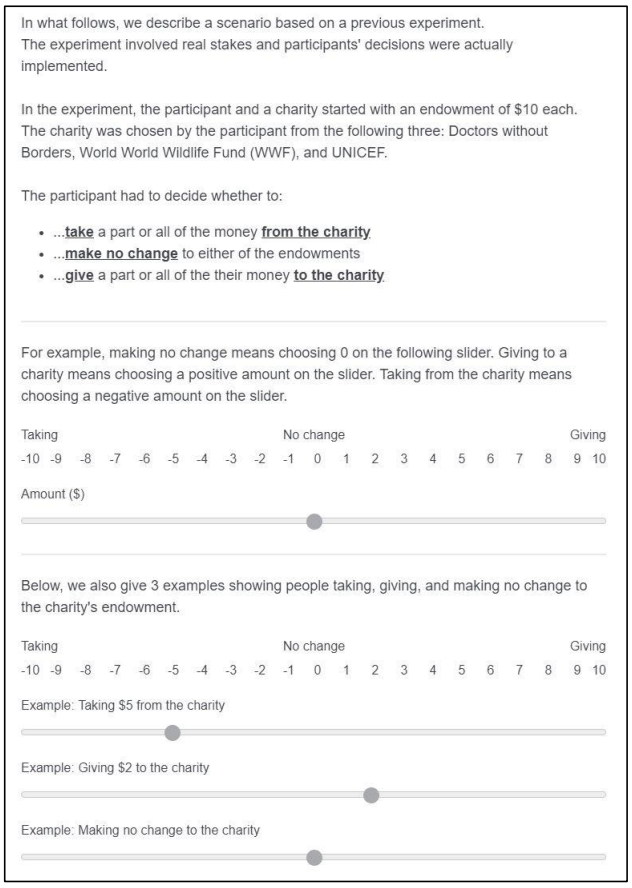


Thank you for choosing to participate in this study! It is important that you **read all of the instructions carefully to maximize your earnings**. At the end of the study, you will be asked to fill out a short questionnaire and demographic questions.

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.

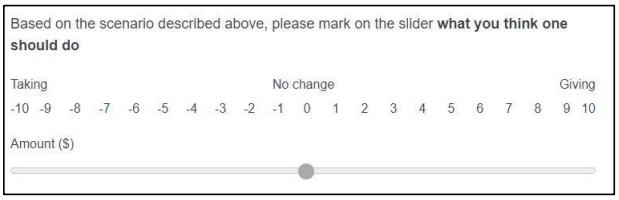
You will be paid $2 for participation and will have the opportunity to earn more based on some of your answers. For example, you will be asked to guess what other participants have previously done or thought should be done. Guessing correctly will earn you an additional $1 for each question that is eligible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

Scenario

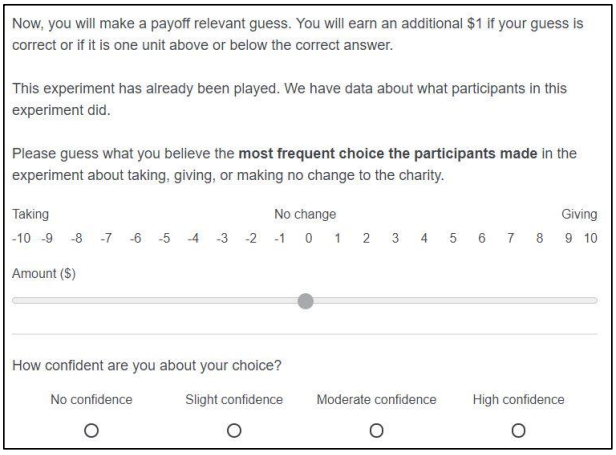

In what follows, we describe a scenario based on a previous experiment.
The experiment involved real stakes and participants' decisions were actually implemented.
In the experiment, the participant and a charity started with an endowment of $10 each. The charity was chosen by the participant from the following three: Doctors without Borders, World World Wildlife Fund (WWF), and UNICEF.
The participant had to decide whether to:
...take a part or all of the money from the charity
...make no change to either of the endowments
...give a part or all of the their money to the charity
For example, making no change means choosing 0 on the following slider. Giving to a charity means choosing a positive amount on the slider. Taking from the charity means choosing a negative amount on the slider.
Taking
No change
Giving
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10
Amount ($)
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

Three Belief Elicitations (Within-Design)

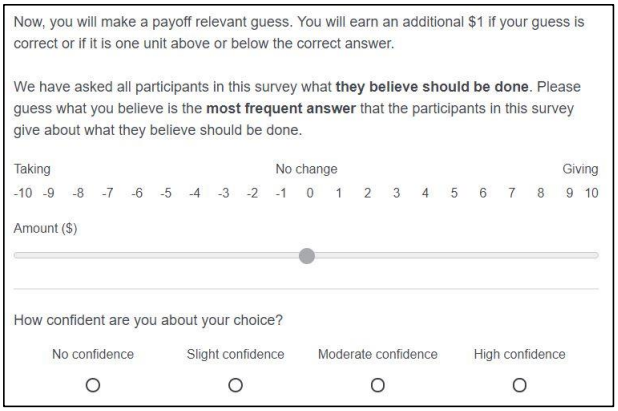
1. *Personal Normative Beliefs (PNB) - Decision*



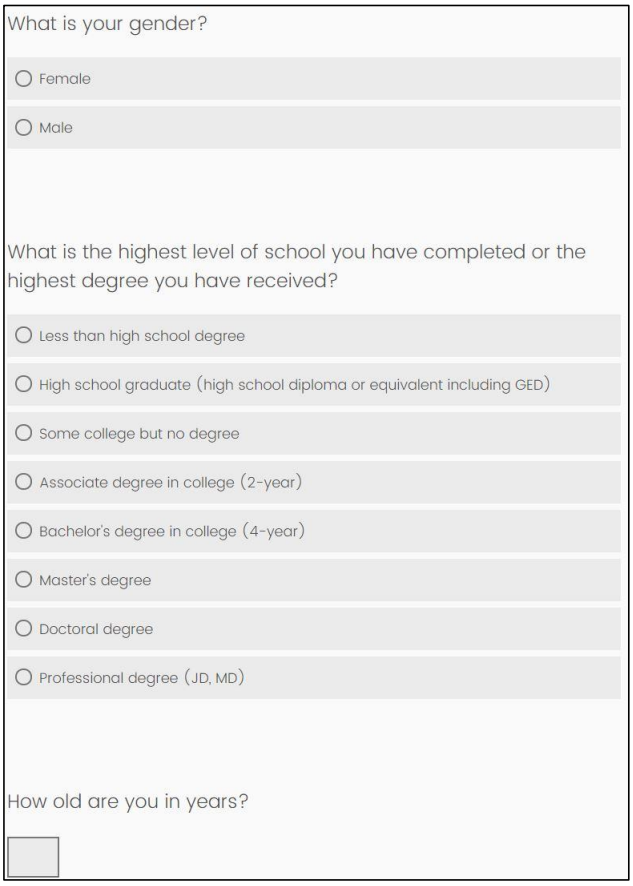
2. *Normative Expectations (NE) - Decision*

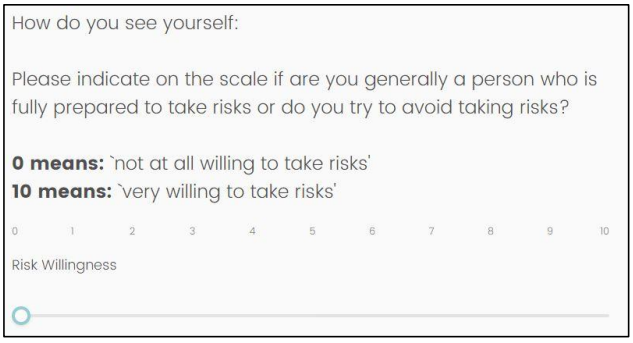


3. *Empirical Expectations (EE) – Decision*



Post-Experimental Questionnaire



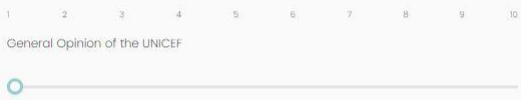


What is your general opinion of charitable organizations?

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

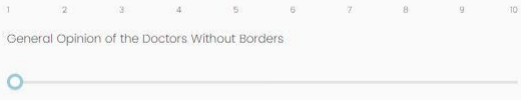
What is your general opinion of the United Nations International Children's Emergency Fund (UNICEF)?

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



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.



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

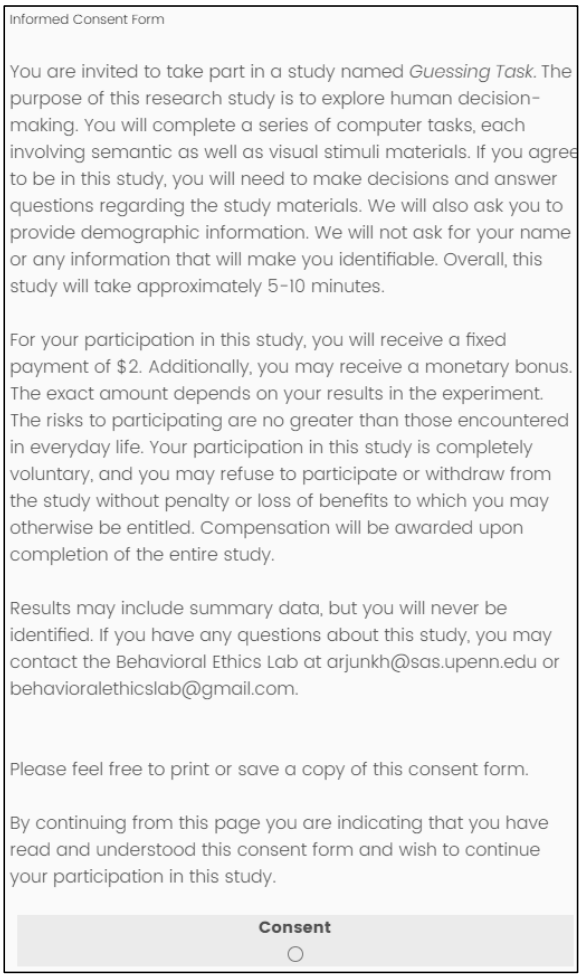
'1' = I do not support the work of the WWF.
'10' = I fully support the work of the WWF.

III.b.1 Screenshots Norm Updating Experiment

To confirm robustness of this, we also run the exact same questionnaire on Amazon Mechanical Turk with $N = 303$ participants. We ensure high quality data collection by utilizing a combination of CAPTCHAs and sophisticated screening questions to avoid pool contamination. We applied the following restrictions to the subject pool: subjects had to be in the United States, and approval rating $> 99\%$. We used online tools to test IPs for low quality respondents. We obtain qualitatively the same results (normative expectations adjust downwards, no changes in personal normative beliefs), which are available upon request.

We can use this initial elicitation of beliefs to gauge the similarity between MTurkers' and Penn students' beliefs (compare Figures **??** and **??**. We find that MTurkers generally hold weaker normative and empirical beliefs than Penn students. Nevertheless, even among MTurkers, a norm of giving (or at least of not-taking) exists: 61% of MTurkers personally believe that one ought to give and 69% believe that this what most other people also believe (the corresponding figures for Penn students are: 72% and 86%). Moreover, 69% expect participants in the experiment to either give or leave the charity endowment untouched (it is 85% among Penn students).

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.

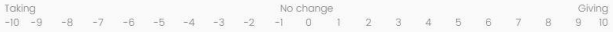
In the experiment, the participant and a charity started with an endowment of $10 each. The charity was chosen by the participant from the following three: Doctors without Borders, World World Wildlife Fund (WWF), and UNICEF.

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.

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



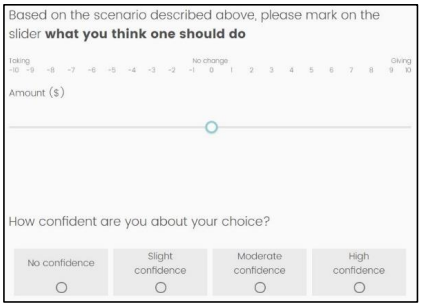
Example: Taking $5 from the charity

Example: Giving $2 to the charity

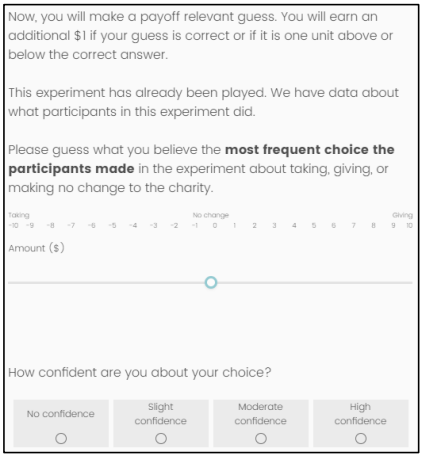
Example: Making no change to the charity

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

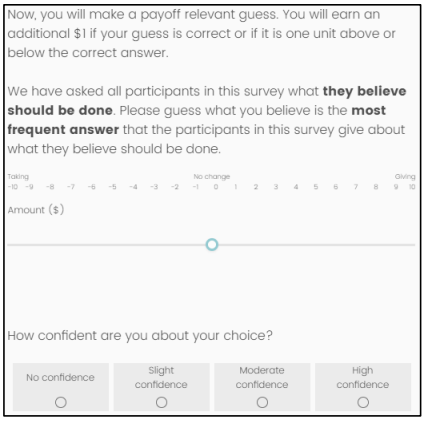
1. *Personal Normative Beliefs (PNB) - Decision*



2. *Normative Expectations (NE) - Decision*



3. *Empirical Expectations (EE) – Decision*



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)

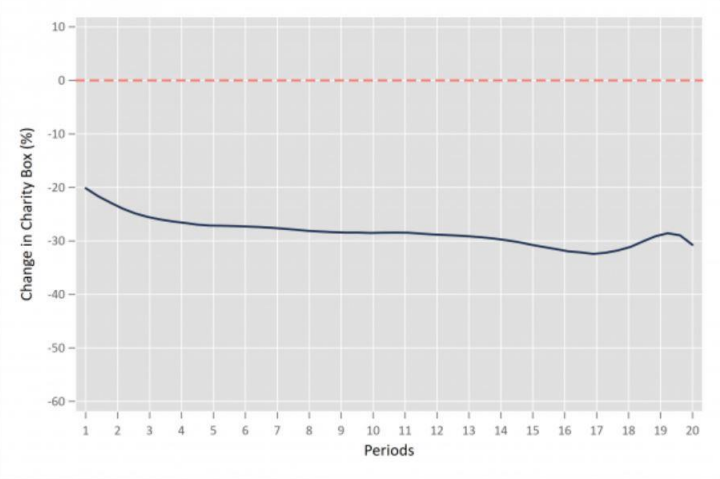
You will observe the *average behavior* of all participants towards the charity box in a particular condition.

- 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

Presented Behavior: NoObservation (Baseline)

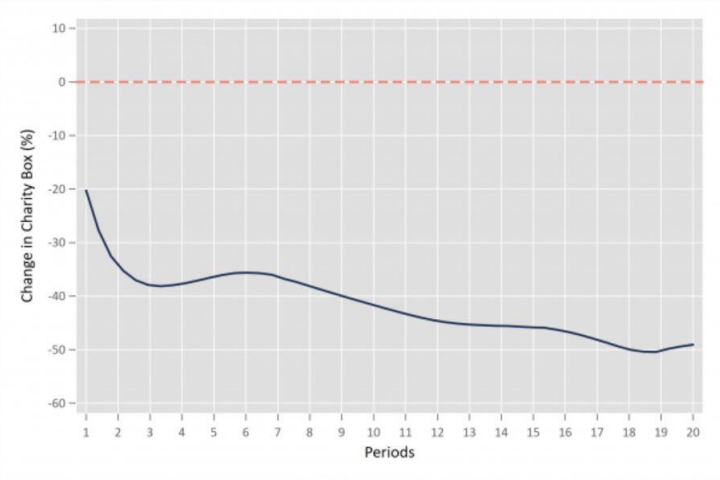
Below is the behavior (blue line) of previous participants towards their charity in the original experiment over the course of 20 rounds. Values below (above) zero indicate that participants took money away from (gave money to) charity. A value of 0 indicates that an equal distribution (= initial starting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



○ Understood

Presented Behavior: Observation (T1)

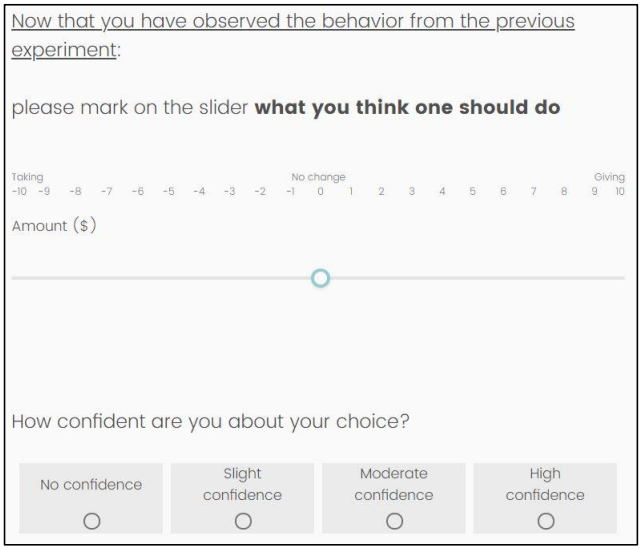
Below is the behavior (blue line) of previous participants towards their charity in the original experiment over the course of 20 rounds. Values below (above) zero indicate that participants took money away from (gave money to) charity. A value of 0 indicates that an equal distribution (= initial starting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



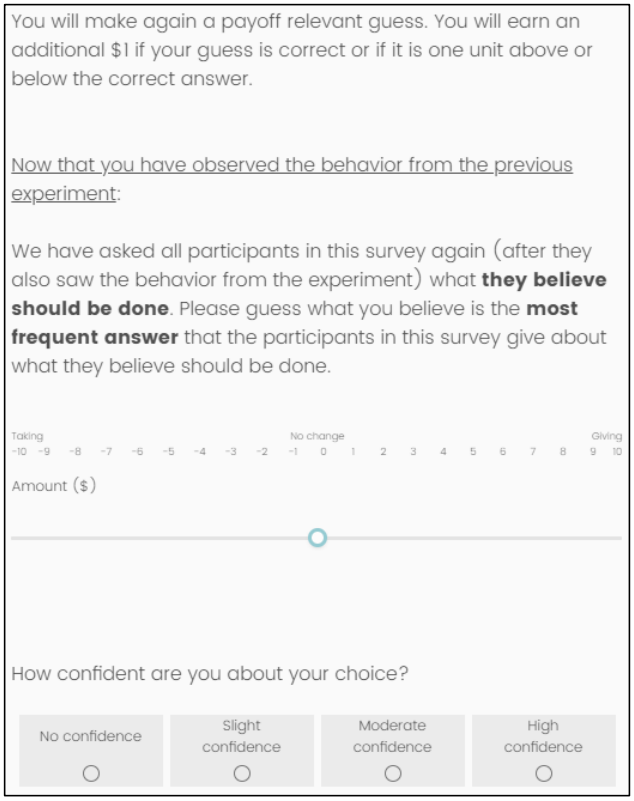
○ Understood

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

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

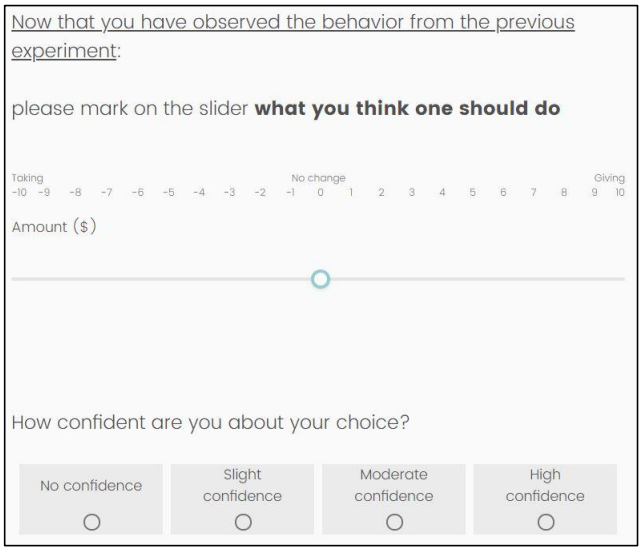


2. *Normative Expectations (NE) - Decision*

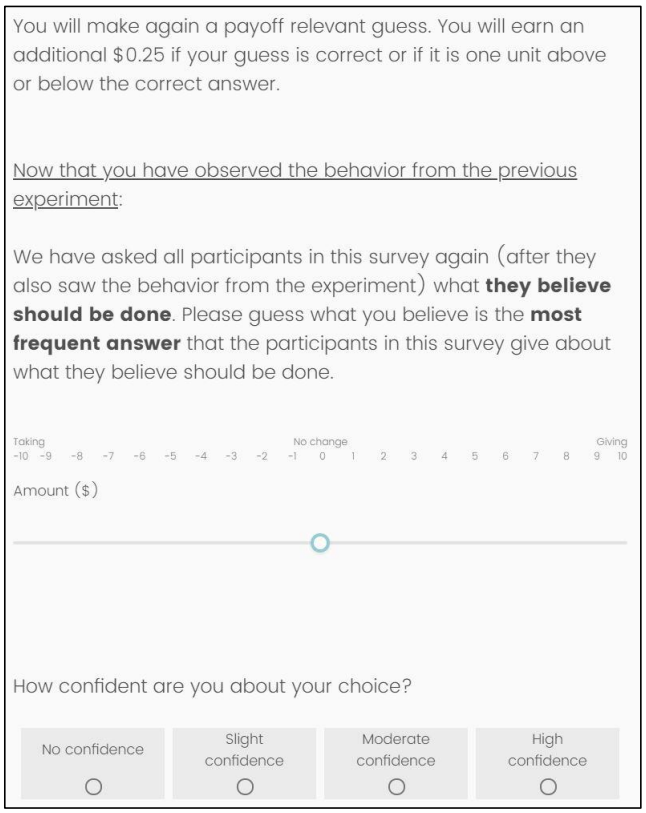


Two Belief Elicitations (Within-Design) – After Observation Stage

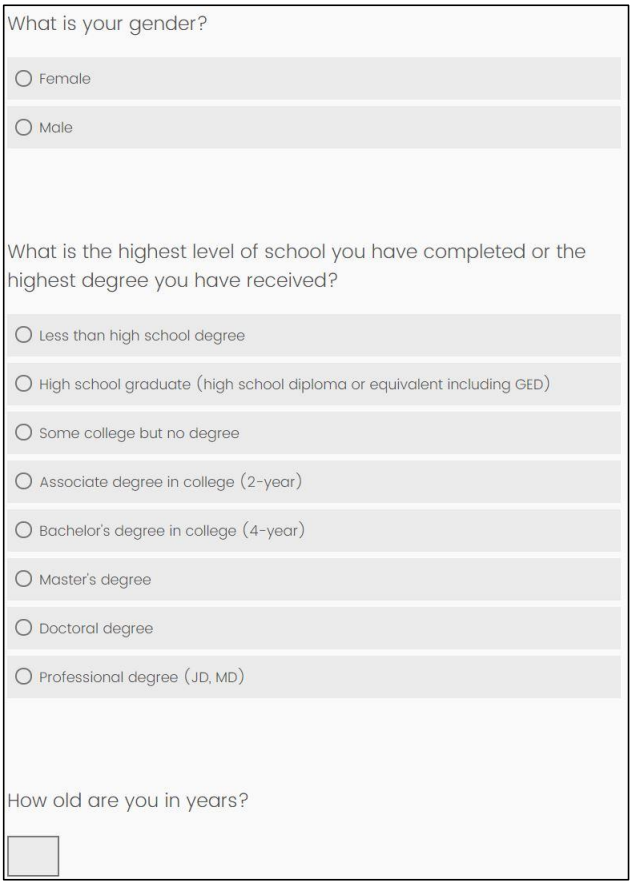
1. *Personal Normative Beliefs (PNB) - Decision*

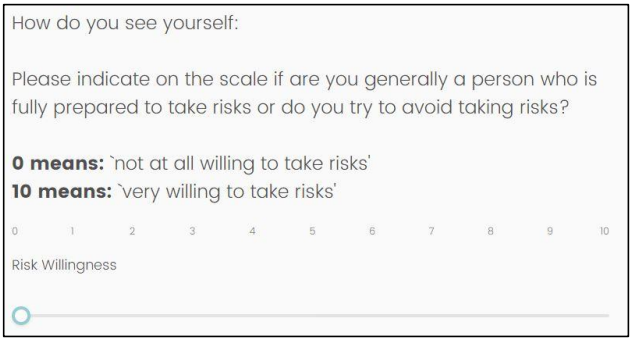


2. *Normative Expectations (NE) - Decision*



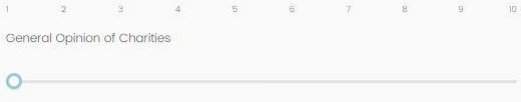
Post-Experimental Questionnaire





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.

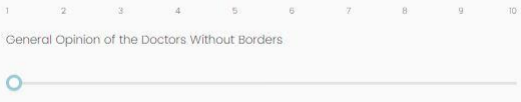


What is your general opinion of the United Nations International Children's Emergency Fund (UNICEF)?

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

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.



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.